Course corrections

If the US government chooses to bail out the nation's vehicle manufacturers, it must ensure that the industry commits to the innovations required for future transportation.

Just months after the US Congress approved $25 billion in loan guarantees for the nation's vehicle manufacturers, the 'Big Three' companies are back in Washington asking for more. The loan guarantees are to help the firms retool their facilities and begin building vehicles with improved fuel efficiency. The latest request from General Motors, Ford and Chrysler is for another $25 billion to help them ride out the current financial crisis.

The reception on Capitol Hill to the bail-out request has so far been frosty, and with good reason. These companies have systematically failed to acknowledge, let alone anticipate, the energy and business trends that are driving an increasingly globalized vehicle market. They have repeatedly fended off fuel-efficiency regulations that might have softened the blow of the recent energy crunch, and instead kept turning out the petrol-hungry sport-utility vehicles that US consumers once embraced, but are now abandoning. This is just one of many mistakes in an outdated business model that has left the Big Three especially vulnerable to the current crisis.

Regardless, General Motors says it will run out of cash in a matter of weeks, and the other two aren't faring much better. As their collapse would only make a bad situation worse, the US government may have no choice but to intervene, just as it felt compelled to bail out Wall Street. This makes it all the more important that any federal rescue plan should be structured not as a short-term monetary fix, but as a chance for US vehicle makers to clear the slate and start anew. As president-elect Barack Obama correctly put it, the government should not be issuing "a bridge loan to nowhere".

In particular, the worst thing policy-makers could do is to raise the bail-out cash by raiding the $25 billion in loan guarantees designated for the manufacturers' green retooling — as some in Congress have suggested. The global transportation sector is responsible for some 13% of the world's greenhouse-gas emissions. Properly implemented, these loans could help turn that situation around.

Even more importantly, vehicle makers and policy-makers alike need to be open to radically new ways of doing business. To paraphrase Albert Einstein, we can't solve problems with the same thinking that created them in the first place. On page 436, for example, *Nature* looks at the prospects for electrification of the transport sector; one clear lesson is that innovation comes in many forms. As companies such as Better Place of Palo Alto, California, and Th!nk of Snarøya, Norway, illustrate, the barriers to new forms of transportation may not lie so much with the technology but with the way it is marketed. By simply leasing batteries and charging customers a monthly fee that includes all electric 'fuel', these companies hope to lower the upfront costs and give consumers the reliability that they want. A similar business model worked for mobile phones, why not for cars?

> "The US government should not be issuing a bridge loan to nowhere."

Whatever happens with that particular approach, electric vehicles increasingly look like a viable part of the solution to the transportation challenge. But getting there will require a multi-pronged approach involving battery manufacturers, vehicle manufacturers, electric utilities and governments. The good news is that utilities and vehicle companies understand this. In fact, some utilities are pondering the idea of jumpstarting the electric car market by ordering vehicles themselves. That is an idea that governments should consider as well.

All of this should be at the fore as the US government ponders investing in the vehicle industry. Government intervention might well be the catalyst for change — but new thinking is going to be required from everyone. ∎

A fruitless campaign

Another protracted fight over genetically modified crops in Africa will be costly and wasteful.

The global food crisis that came to the fore last spring may have been overshadowed by the global financial crisis that erupted this autumn, but it has certainly not been solved. That is one reason why many governments and philanthropic foundations are now looking to agricultural biotechnology to improve future food production. Despite the virulent opposition to genetically modified (GM) crops in some quarters, many believe that progress in areas such as drought-tolerant or nutritionally fortified plants could make a big difference in many of the poorest countries.

Indeed, environmentalists, policy-makers, scientists and industry representatives have been meeting both formally and informally over the past few years — first to establish a degree of common ground, and then to approach the trickier business of bridging some of their differences on the role of GM technology in agriculture.

A prime example is the work of the African Union's High-Level Panel on Modern Biotechnology, which was charged with charting a way forward in what have become known as Africa's GM wars. For well over a decade, companies such as Monsanto have sought to create African markets for GM crops such as insect-resistant *Bt* cotton, while against them have stood European environmental groups and not a few African political leaders, for whom multinational businesses evoke the spectre of colonialism. The two sides have waged a war in parliaments, in the media and even on the streets.

Fed up, the African Union eventually brought together a group

of key individuals and institutions who might otherwise be talking to each other through a megaphone. The group included Tewolde Egziabher, head of Ethiopia's Environmental Protection Authority, who is a leading environmental campaigner and a vocal critic of multinationals in developing countries. Sat next to him was Calestous Juma, a professor of international development at Harvard University and a passionate proponent of technology's role in economic development. And next to him was Cheick Modibo Diarra, chairman of Microsoft in Africa.

The group eventually came to a consensus that Africa's nations cannot afford to do without new technologies in agriculture — but that all new technologies would need appropriate safeguards to protect human health and the environment. This seemingly obvious statement was, in fact, a rare example of successful collaboration between multinationals and environmentalists.

The fragility of that consensus is illustrated by the fate of a much larger initiative, the International Assessment of Agricultural Knowledge, Science and Technology for Development. That effort attempted to forge a similar consensus among the major players in world agriculture, but fell apart in January when industry representatives chose to walk away from the table (see *Nature* **451,** 223–224; 2008). They felt unable to sign a document that did not list biotechnology as a high enough priority.

From the other side, meanwhile, GM opponents are trying to rekindle the controversy. A new opposition campaign — www.bangmfood.org — was endorsed in the November issue of *The Ecologist* magazine, an influential voice in the global environmental movement.

In that context, the magnitude of the African Union panel's achievement is clear — as are the challenges it still faces. Its report, *Freedom to Innovate: Biotechnology in Africa's Development*, has not yet officially seen the light of day, even though it was published more than a year ago. Ordinarily, a document from the African Union would be expected to be harsh in its criticism of multinational industry. As this report is more measured, senior officials in the African Union's Commission based in Addis Ababa are nervous about releasing it.

> **"Africa's nations cannot afford to do without new technologies in agriculture."**

Happily, the report is in wide circulation and freely available on the Internet (www.nepadst.org/doclibrary/pdfs/biotech_africarep_2007.pdf). But the African Union should have the courage of its convictions and give the report its formal endorsement. Indeed, it should use it as a model for ongoing attempts to address the food crisis. Both the successful negotiations of the African Union panel and the failure of the international assessment show that there is no alternative to a grown-up discussion with all parties in the same room. ■

Friendly rivalry

The spirit of collaboration in the race to define the LHC's successor sets an example for large projects.

The future for high-energy physics is decidedly mixed. On the one hand, physicists are eagerly awaiting the insights into the Universe promised by the Large Hadron Collider (LHC) at CERN, the European particle-physics laboratory near Geneva. But as governments shift their priorities to societal problems, such as climate change, energy, health and the environment, the field as a whole must also face up to the fact that it will be increasingly difficult to secure funds for pure science.

Given this financial uncertainty, it is important that the high-energy physics community does all it can to reduce any internal divisions and to strengthen its external coherence. That is why a new collaboration over what should come after the LHC is to be greeted with enthusiasm. As a new world-class accelerator requires a decade or more of technology development and coalition-building, physicists are already laying their plans. But until the LHC starts providing data, there will be uncertainty over the energy levels at which its successor should operate. So two competing concepts, backed by rival efforts, have been put forward.

On one side is the International Linear Collider (ILC), which would use electron–positron collisions to pick up where the proton–proton collisions of the LHC leave off (see *Nature* **451,** 108; 2008). The ILC planners are setting their sights on comparatively low energies — about 500 gigaelectronvolts per beam — a goal that could be achieved fairly easily with technology that is now maturing. The United States, seeking a return to pre-eminence at the high-energy frontier, has been a driving political force for hosting the ILC. On the other side is the Compact Linear Collider (CLIC) being pushed by CERN. This could conceivably run at energies as high as 3 teraelectronvolts — but the technology to achieve that lies much farther in the future.

The potential for destructive rivalry was real. Yet late last month, leaders of the two efforts formally agreed to collaborate as much as is practicable. Seven joint working groups have been established, covering common efforts ranging from detectors to beam-delivery systems. In a first, the working groups presented joint talks at CLIC and ILC workshops during the past few weeks. One group is trying to put costs and schedules into common terms so that decision makers will find it easier to compare the two when it comes to choosing between the technologies. Another group will save money by applying a CLIC simulation to an ILC design for an accelerator component known as a damping ring. The two rivals are closer than they have ever been, and yet research and development on the two underlying accelerator technologies will continue apace with a healthy spirit of competition.

The *rapprochement* was not entirely brought about by the current financial crisis — Barry Barish, director of the ILC global design effort, says that talks began well before December 2007, when the US Congress passed a budget that derailed much of the US ILC effort. But the global economic meltdown has certainly made cooperation imperative — not least because it has pushed the next big accelerator even farther into the future. The result is that the ILC and CLIC are setting an example that other large scientific endeavours would do well to emulate. ■

RESEARCH HIGHLIGHTS

Breakin' the law

Science doi:10.1126/science.1161405 (2008)

The controversial 'broken windows' thesis holds that small signs of crime, such as smashed glass and graffiti, can lead to more of the same, or worse. This is the idea behind 'zero tolerance' policing. But strong, supportive evidence for the theory has been lacking.

Kees Keizer and his colleagues at the University of Groningen in the Netherlands covertly observed fellow Groningers who were given the opportunity to steal an envelope that obviously contained a €5 note from a postbox. When the postbox was clean and tidy 13% took the bait; by contrast, 27% stole from a graffitied postbox and 25% from one with litter around it.

Other tests showed that people are more likely to litter in the presence of graffiti or abandoned shopping trollies, and after hearing the crackle of illegal fireworks.

SPACE SCIENCE
Colour test

Icarus doi:10.1016/j.icarus.2008.10.006 (2008)

Why is reddish Pluto a different colour from its three greyish moons, Charon, Nix and Hydra? Because the three are regularly resurfaced with loose material thrown up in the weak gravity of Nix and Hydra by impacts with small outer-Solar-System objects. Meanwhile, material ejected fast enough to travel as far as Pluto tends to be moving with sufficient speed to escape the gravity of the Pluto system altogether, according to Alan Stern of the Lunar and Planetary Institute in Houston, Texas. Pluto also repaints itself with an annual frost, and is partly shielded by Charon.

By assuming that the three moons are coated in the same stuff, Stern predicts that Nix and Hydra are both about 50 kilometres in diameter. NASA's New Horizons spacecraft, which Stern is in charge of, will find out for sure in 2015.

BIOLOGY
Mile-high strategies

J. Anim. Ecol. doi: 10.1111/j.1365-2656.2008.01491.x (2008)

A sparrow found in the northern reaches of the Rocky Mountains of North America shows a tradeoff between lifespan and reproductive period according to how high up it lives, and thus the weather conditions and food available to it.

Populations of dark-eyed juncos (*Junco hyemalis*; pictured right) in Canada's Jasper National Park are regularly found living as much as 2,000 metres above sea level. Heather Bears at the University of British Columbia and her colleagues monitored

the songbirds at four 2,000-metre sites and at four 1,000-metre-sites. They found that juncos at the higher elevation lived longer and produced 55–61% fewer offspring per year than those living lower down. When these high-altitude birds did reproduce, however, they invested more heavily in each offspring; their chicks were 15–20% more likely to survive into adulthood and weighed, on average, 11% more at 25 days old.

VIROLOGY
Inside knowledge

Nature Med. doi:10.1038/nm.1885 (2008)

During replication, many viruses trigger certain lipid molecules that are normally found only on the inner layer of the cell membrane to spread into the outer layer. An antibody that looks for these lipids on cell surfaces can selectively pick off infected cells.

Philip Thorpe of the University of Texas Southwestern Medical Center in Dallas and his colleagues have shown this with the antibody bavituximab, which seeks the lipid phosphatidylserine. They cured guinea pigs

with lethal Pichinde virus infections and mice harbouring deadly cytomegalovirus. The former virus is a model for the Lassa fever virus, a potential bioterror agent.

THEORETICAL PHYSICS
Eyeing entanglement

Phys. Rev. A **78**, 052110 (2008)

In 1964, John Bell suggested that if Bob saw one of a pair of photons and Alice saw the other, the duo would measure a shared quantum state only if the photons were entangled. But Bob and Alice needn't have carried man-made photon detectors, according to calculations by Nicolas Brunner and his co-workers at the University of Geneva in Switzerland: by adding a few more photons, their eyes would have been good enough.

A healthy human eye cannot spot a single photon, but can usually pick up fewer than ten. This threshold should not hinder a person's ability to detect entanglement in a similar set-up to Bell's experiment in which several photon pairs are emitted at once. Brunner suggests four observers for the task — two Alices and two Bobs — so that the observers can agree on whether they really saw such low light intensities.

ONCOLOGY
Odd ones out

Science doi:10.1126/science.1160165 (2008)

By tracking the expression and location of almost 1,000 proteins in living cells, researchers have found new clues to how some cancer cells resist the chemotherapy drug camptothecin.

Camptothecin targets a DNA-binding protein called TOP1. Ariel Cohen of the

Weizmann Institute of Science in Rehovot, Israel, and his colleagues labelled proteins in human lung cancer cells with different fluorescent tags, and then tracked the proteins for two days after camptothecin treatment.

About one out of every 10,000 cells survived the treatment. The reason seems to lie in the behaviour of two of the labelled proteins, DDX5 and RFC1, levels of which increased in surviving cells but decreased in those that died.

SYSTEMS BIOLOGY
Radical methods

Cell **135,** 679–690 (2008)

In one of the first demonstrations of the practical value of systems biology, James Collins and his colleagues at Boston University in Massachusetts have shown how the aminoglycoside class of antibiotics works. Although it was known that these compounds disrupt ribosomes — the cell's protein-making machinery — and cause proteins to be wrongly made, exactly how this killed bacteria was unclear.

Using DNA microarrays, Collins and his team mapped out changes in the gene-expression patterns of *Escherichia coli* treated with one or other of the aminoglycosides gentamicin and kanamycin. This allowed the researchers to pin down the gene networks that the drugs influenced. From these data and studies with bacterial mutants, they worked out that the movement of certain defective membrane proteins creates an electrochemical gradient across cell membranes that ultimately spawns harmful hydroxyl free radicals.

MOLECULAR BIOLOGY
Cancer 101

Science doi:10.1126/science.1165395 (2008)

An RNA molecule called microRNA-101 inhibits the production and function of a protein that is found at high levels in many types of aggressive tumour.

Pinning this down first required Arul Chinnaiyan of the University of Michigan Medical School in Ann Arbor and his co-workers to predict the various microRNA molecules with the right shape to target the protein, EZH2. Four computer programs listed 29 possibilities in total, but only two of these appeared in all four readouts. Of these, only miRNA-101 reduced the amount of EZH2 when precursors of this RNA were added to cancer cells.

The researchers also found that one or both of two genes that encode miRNA-101 had been lost in two-thirds of spreading prostate cancers, unleashing overexpression of EZH2.

PHYSICAL CHEMISTRY
Electrons with a twist

Phys. Rev. Lett. **101,** 178301 (2008)

The direction of an electron's spin is sufficient to determine the decomposition rate of the two mirror-image forms of a molecule.

Richard Rosenberg of Argonne National Laboratory in Illinois and his colleagues drew this conclusion after firing X-rays at a magnetic iron–nickel alloy coated with thin layers of 2-butanol. This freed electrons from the alloy, and their spins aligned with its magnetic field.

With the alloy magnetized in one direction, the electrons tended to increase the rate that right-handed 2-butanol fell apart in response to light by about 10%; the opposite field caused a similar effect in left-handed molecules. The two forms of the molecule are shown below.
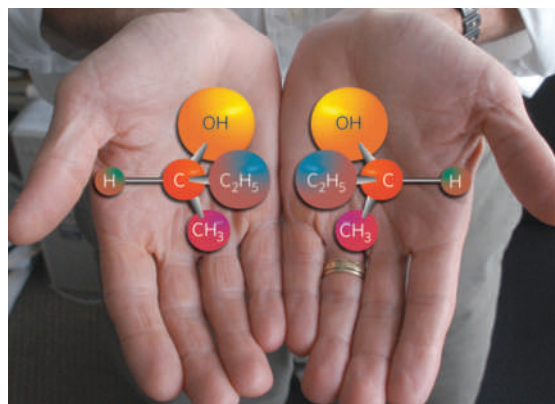
ATMOSPHERIC SCIENCES
Methane mutterings

Geophys. Res. Lett. doi: 10.1029/2008GL036037 (2008)

After almost a decade of stability, the amount of methane in the atmosphere has been growing since the start of 2007, according to measurements taken by two global monitoring networks, AGAGE and CSIRO.

The findings are cause for concern because methane is many times more powerful than carbon dioxide as a greenhouse gas, and has so far accounted for about a fifth of the human contribution to climate change. Matthew Rigby of the Massachusetts Institute of Technology in Cambridge and their co-workers write that, by late 2007, the proportion of methane in the atmosphere was rising by 10 parts per billion per year at all monitoring stations around the world.

High bacterial methane emissions from wetlands in an unusually warm Siberia seem to have played some part in the northern hemisphere increase. The authors also suspect that an unproven drop in hydroxyl free radicals could be driving the trend.

JOURNAL CLUB

**William B. McKinnon
Washington University, St Louis, Missouri**

A planetary scientist has big hopes for a little world.

Right now, the most exciting object in the Solar System is Saturn's diminutive moon Enceladus. Its deformed south polar region emits copious amounts of heat along the length of several young, active ridges and fractures, as well as plumes of tiny ice particles, water vapour and other chemicals.

The Cassini spacecraft — equipped with plume-gas and particle analysers and clever imaging gadgetry — is currently in the neighbourhood. Seizing this opportunity, Gabriel Tobie of the University of Nantes in France and his colleagues have incorporated some of its recent measurements into theoretical models of tidal heat production on Enceladus (G. Tobie *et al. Icarus* **196,** 642–652; 2008). Only the ebb and flow of tides could properly account for such prodigious geological activity on an icy moon that measures just 500 kilometres in diameter.

The authors start with the generally accepted idea that Enceladus has differentiated into a rock core and an icy mantle. They then show that the size of the tidal motion of the mantle is inadequate to generate the observed thermal emission, so there must be a fluid ocean sandwiched between the two solid layers. This is no great surprise, but Tobie *et al.* go further, showing that even if the mantle is made soft and deformable over the southern polar region (as the ice would be if it were relatively warm), a sandwiched, liquid ocean must reach at least as far as around the entire southern hemisphere.

The team imagines that, below Enceladus's south pole, tidal heating concentrates in warm, upwelling, convectively mobile ice. This, in turn, causes the cold, brittle surface layer to rupture — and the exposed warm ice sublimates, releasing trapped gases. It is a compelling picture, and one that promises to help unlock the internal activity of other icy satellites.

Discuss this paper at **http://blogs. nature.com/nature/journalclub**

NEWS

Beamline bonanza for Japanese researchers

A proton-accelerator complex will incorporate one of the world's most intense pulsed neutron sources.

TOKAIMURA, JAPAN

Staff at the new Japan Proton Accelerator Research Complex (J-PARC), in the seaside village of Tokaimura in Ibaraki, like to use familiar analogies to introduce the facility. The 330-metre-long building that houses the linear proton accelerator — the engine behind the complex — is similar in size to Tokyo's train station. And the main 50-gigaelectronvolt synchrotron, with its three flattened sides and a diameter of 500 metres, resembles an *onigiri* — the rice ball found in every convenience store throughout Japan.

J-PARC will support several fields in which Japan has excelled, and if organizers have their way, Japan will stay at the forefront of them. Starting next month, the facility will accelerate and smash protons into a variety of targets to produce beams of subatomic particles such as neutrons, muons, neutrinos and kaons. It promises to provide its users — including a large proportion of international scientists — with new ways to image molecules, probe neutrino physics and more.

Most accelerators are built either to study nuclear physics or fundamental particle physics, or for materials or medical use. J-PARC is designed to do all of that, says director Shoji Nagamiya, in part because it sprang from two earlier proposals that merged. The Tsukuba-based High Energy Accelerator Research Organization was looking to build a 50-gigaelectronvolt ring for neutrino and kaon projects; the Japan Atomic Research Energy Institute, which was later reorganized as the Japan Atomic Energy Agency, wanted a neutron beam. The idea to share an accelerator, which came about in 1998, made sense, but the union was consumed by red tape. The institutes finally signed off on what would become a ¥152-billion (US$1.6-billion) collaboration in 2001. "J-PARC is like an indispensable, high-quality single crystal finally grown after a tremendous struggle," says Kazuyoshi Yamada, a materials researcher at Tohoku University in Sendai who started preparing samples for it two years ago.

At the complex's materials and life-science facility, research kicks off in earnest next month when seven neutron beamlines start running. These will be used to precisely image biological molecules and inorganic crystals, among other things. For instance, Nobuo Niimura of Ibaraki University will lead a protein-structure project; researchers from Hitachi will analyse lithium-based materials used for plug-in electric cars (see page 436); and yet others will study the behaviour of magnetic fields inside superconductors.

J-PARC's neutron beams will be 1 megawatt — intense enough to dramatically speed up the rate of discovery. Earlier neutron studies, using neutrons produced in nuclear reactors, led to only five or so protein structures being completed in a year, says Niimura. At the new complex, this number will jump to 100. Thomas Mason, director of the Spallation Neutron Source at the Oak Ridge National Laboratory in Tennessee (see 'Neutron sources around the world'), says that the two facilities are "a new threshold in performance". In 2009, J-PARC will also start running experiments with the world's most intense pulsed muon beam, exceeding by 30-fold the Rutherford Appleton Laboratory's facility in Didcot, UK.

Meanwhile, the largest and costliest part of the facility — the 50-gigaelectronvolt synchrotron — will be devoted to studies of neutrinos and kaons. A team of some 400 researchers will race to observe for the first time the transition of one variety of neutrinos, called muon neutrinos, into another called electron neutrinos. The observation could offer clues about the range of neutrino masses and about CP violation, a concept key to understanding why matter dominates over antimatter in the Universe.

Groups in China and France are pursuing a similar goal, but with less-energetic neutrinos generated by nuclear reactors. The Fermi National Accelerator Laboratory in Batavia, Illinois, will have a more powerful neutrino source when its NOvA gets up and running after 2014.

Neutron sources around the world

Japan's Proton Accelerator Research Complex is not the only powerful neutron source to fire up these days.

In Tennessee, the Spallation Neutron Source at the Oak Ridge National Laboratory is already the world's most intense pulsed neutron source at 630 kilowatts, with a projected ramp-up to 1.4 megawatts in the next two years and a projected top-out at 4 megawatts. Two science instruments are coming online by the end of the year, bringing the total to nine operating instruments.

Not to be left behind, the United Kingdom is refurbishing and adding to its 25-year-old ISIS neutron source. This 160-kilowatt facility already hosts 29 operating instruments, and in August it added a second, £146-million (US$218-million) proton target station with seven new instruments. In France, the continuous, 58-megawatt neutron source at the Institut Laue-Langevin now has €160 million (US$202 million) to upgrade the facility's 40-plus instruments.

As for the future, China could stake its claim with the projected 2015 opening of the 100-kilowatt China Spallation Neutron Source. The Dongguan-based facility is to be ISIS-like in design, with an estimated 1.4-billion-renminbi (US$205-million) price tag.

But leading the field could be Europe's proposed 5-megawatt European Spallation Source. If project leaders can pick a site for it in the next two months, the €1.5-billion neutron source could come online in 2022. **Ashley Yeager**

D. CYRANOSKI

Japan hopes its 50-gigaelectronvolt synchrotron will boost the nation's standing in medical imaging and particle physics.

To create the beam of muon neutrinos, J-PARC will send a proton beam at a graphite target. Unlike conventional head-on collisions, J-PARC's proton beam will be adjustable to allow collisions at an angle. The angled collision will lower the average energy, but more of the resulting neutrinos will be in the range likely to oscillate into electron neutrinos on their underground trip to the SuperKamiokande detector, 295 kilometres away.

When J-PARC's big synchrotron isn't doing neutrino physics, it will focus on kaons and hypernuclear experiments. For example, scientists will look for evidence that neutral kaons decay into neutral pions plus a neutrino and an antineutrino.

About half of the $10 million needed for the decay experiment, and $8 million needed for other kaon experiments, still await funding. But this year's Nobel prizes may help to give the field a boost. In 2002, the awarding of the Nobel to Masatoshi Koshiba, then at the University of Tokyo, for neutrino work meant that Japan ploughed another ¥16 billion into J-PARC's neutrino project, accelerating its opening. This year, the Nobel physics prize was shared by the High Energy Accelerator Research Organization's Makoto Kobayashi and Kyoto University's Toshihide Maskawa for their work in explaining the dominance of matter over antimatter in the Universe — and some think that might boost funding for the kaon experiments at J-PARC. "It was," says Nagamiya, "very good timing." ■
David Cyranoski

Astronomers unveil wish list

European scientists and funding agencies have launched an ambitious plan to prioritize the astronomy projects they believe should be supported over the next 20 years.

The ASTRONET Infrastructure Roadmap was released on 25 November, timed to coincide with the opening of the European Space Agency's (ESA's) ministerial meeting in The Hague, the Netherlands, where, as *Nature* went to press, ESA member states were due to thrash out how planned space projects will be funded over the coming years. Several large ESA projects — including a suite of Earth-observing satellites called Kopernikus and the proposed ExoMars rover — are already facing funding difficulties (see *Nature* 455, 840–841 & 1013; 2008).

ASTRONET was set up in 2005 to deliver a concerted vision from the European astronomy community. It echoes a process that has been undertaken five times in the United States — the National Research Council's 'decadal survey' of astronomy and astrophysics.

Speaking with a single voice has helped US space scientists to get big projects such as the Hubble and Spitzer telescopes off the ground. Astronomer Joseph Taylor at Princeton University, who co-chaired the last decadal survey in 2001, thinks Europe will benefit from having a consolidated approach. "Over the past 50 years it's been a very useful way of prioritizing our national efforts," he says.

The European version collates views from 28 countries, as well as ESA and the European Southern Observatory (ESO). The process is a first for European astronomy, but could happen every five to ten years, says Michael Bode at Liverpool John Moores University, UK, and leader of ASTRONET.

ASTRONET's top priority is the European Extremely Large Telescope, a 42-metre telescope operating in the visible and infrared wavelengths that would be the biggest yet built. The roadmap says that construction should begin at the ESO in 2010, backed by €1 billion (US$1.3 billion) in funding that has yet to be committed. The Square Kilometre Array, a radio telescope aiming to probe the early Universe, should be the next priority, and would involve a global consortium of partners.

ESA has its own 'Cosmic Vision' for space-based missions for 2015–25, and there is inevitable overlap with the ASTRONET plan, says Bode. ESA missions favoured by the roadmap include the gravitational-wave observatory LISA — a joint project with NASA — and a probe called Gaia that will map the position and velocity of a billion stars in our Galaxy.

But the roadmap might be bad news for some facilities, because supporting big projects will mean savings must be made elsewhere. ASTRONET is already reviewing the roles of some smaller, two- to four-metre European telescopes in an attempt to eliminate research overlap, for example. "We think it is important to rationalize the observational set-up in Europe," says Jean-Marie Hameury, ASTRONET coordinator and deputy director of the National Institute of Space Science at the CNRS, France's basic-research agency in Paris. A similar review of eight-metre telescopes should follow, he adds.

One such telescope, the 2.6-metre Nordic Optical Telescope on La Palma in the Canary Islands, is run by ASTRONET board chair Johannes Andersen. He points out that the roadmap's recommendation to integrate space- and ground-based observations — something not currently done at a European level — should produce more science overall.

ASTRONET's plan requires European nations to collectively spend 20% more than the €2 billion they spend annually on astronomy. Securing that funding will be "the toughest part of the job", admits Hameury. ■
Katharine Sanderson

ESO
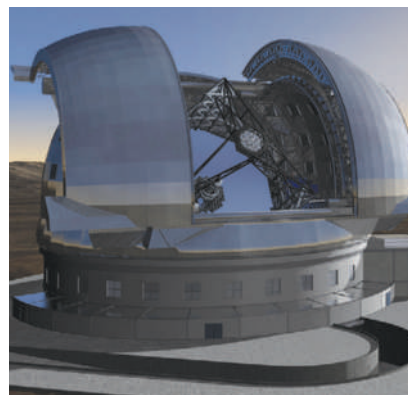
ASTRONET is enthusiastic about the European Extremely Large Telescope.

SPECIAL REPORT

Obama's win refreshes key climate talks

As all eyes turn to Poland for the start of the United Nations meeting next week, **Jeff Tollefson** looks at what progress is likely to be made.

The United Nations Climate Change Conference that begins in Poznań, Poland, on 1 December will in some ways mark the end of an era. The United States' long-standing opposition to climate regulation is vanishing, offering new opportunities for cooperation with its allies in Europe and beyond.

But to some extent, international climate negotiators will remain in limbo until 20 January 2009, when US President-elect Barack Obama enters the White House. Obama has advocated forceful domestic action on global warming and re-engagement with the international community.

"There is a lot of hope and a lot of optimism," says Rob Bradley, who heads international climate policy at the World Resources Institute, an environmental think tank in Washington DC. "A lot of countries will be willing to give the benefit of the doubt to the new administration — but they are all very aware that in Poznań they will be talking to the old administration."

The meeting will bring together representatives from some 192 countries in an ongoing effort to craft a global-warming accord to succeed the Kyoto Protocol, which was adopted in 1997 and expires in 2012. UN officials hope to reach a successor agreement in Copenhagen next year to leave time for implementation and ratification. Yet many think that goal is too ambitious, especially at a time when world leaders are worried about the global economy.

Poland has cited economic reasons in trying to build a coalition to block a European Union rule that, among other things, would require full auctioning of carbon allowances in 2013. But Saleemul Huq, of the International Institute for Environment and Development in London, thinks that opposition is now waning. He says the European Council might even move ahead with the rule as early as 12 December, the last day of the Poznań conference. The US re-engagement will only help, he says. "People will have a more rosy outlook in terms of being able to achieve something," says Huq, "and that will probably bend the European position in a more positive direction."

The United States, which is responsible for most of the greenhouse gases in the atmosphere today, has been largely out of international climate negotiations since President George W. Bush rejected the Kyoto agreement in 2001. If the country re-engages with mandatory emissions limits, it will eliminate an easy excuse for China — now the world's largest emitter — and other rapidly developing nations to remain on the sidelines.

But first the United States must get its own house in order. On 18 November, Obama made clear that he intends to be taken seriously on the international climate stage. In a video-taped address to a governors' climate summit, he reiterated campaign pledges to set up a federal cap-and-trade system for greenhouse gases similar to the European trading scheme, with reductions to 1990 levels by 2020 and 80% below that by 2050.

He also said he had asked any representatives from Congress to report back from Poznań on the talks there — quashing any speculation that he might send his own team of observers. Several other members — including California Democrat Barbara Boxer, who chairs the Senate committee with jurisdiction over global-warming regulation — will send staff.

As it stands, the US delegation will once again be headed by Harlan Watson and Paula Dobriansky, of the State Department, and possibly Jim Connaughton, head of the White House Council on Environmental Quality. Although the trio are likely to lose their posts in the new administration, permanent mid-level staff can continue negotiating countless technical details under Obama. These might include working out how to make emissions reductions among all countries "measurable, reportable and verifiable", and how to create tools and institutions to track factors such as deforestation and compensating countries that preserve their forests.

Deforestation will be among the key issues on the table in Poznań. Last year at a UN climate conference in Bali, international negotiators agreed to include in the next agreement

> **"It's like a ladder: on one side is the United States and on the other is the Chinese."**

reducing deforestation, which releases carbon stored in plants and soils. Papua New Guinea, Costa Rica and other tropical countries have banded together to push for some kind of resolution, seeing the potential for new streams of money flowing into forest-conservation projects from industrialized nations seeking to offset emissions.

Asking the right questions

Negotiators seem to be coalescing around the idea that each country will need to establish national baselines for deforestation, and those that can reduce deforestation below those levels would be eligible for funding. "I think overall we're past the questions of can we measure deforestation, and can we measure reductions in it," says Douglas Boucher, who heads the tropical forests and climate initiative at the Union of Concerned Scientists advocacy group from Washington DC. "It's a question of how we do it."

However, the question of how to pay won't be settled in Poznań — in part because that issue is tied up with broader debates about how and to what extent industrialized nations should provide aid to the developing world.

Poznań could, however, see the launch of a new adaptation fund, as agreed to last year in Bali. The idea is to tap money being collected through a 2% levy on the purchases under the Clean Development Mechanism, which allows developed countries to offset their emissions by funding clean-energy projects or otherwise reducing emissions in developing countries. Funds would be used to help poor countries adapt to a warmer climate — for example,

Action by the United States on greenhouse emissions would put more pressure on countries such as China and Poland to follow suit.

by developing drought-resistant crops or preparing for higher sea levels.

Some, like Bradley, see that as only a first step. He says detailed assessments need to be worked out at the country level before these funds begin to flow, and even then the amount of money is not remotely on the scale needed. The fund could hold from US$270 million to $600 million by 2012, according to World Bank projections. By contrast, the United Nations has called for industrialized nations to provide around $86 billion annually by 2015 to help developing countries cope with a warmer world.

UN climate chief Yvo de Boer acknowledges that progress will be slow in Poznań, but says that is to be expected. He hopes delegates will be able to get their first look at negotiating text and begin fleshing out the details. "I don't think that every meeting needs to be spectacular or can be spectacular," he says. "In fact, Poznań is a meeting at which people just need to get on with their work and move the process forwards."

Political realignment in the United States might open new doors, but it could take weeks or even months for the Obama administration to create its negotiating team and get organized. De Boer says the EU is thinking of proposing another ministerial-level meeting before Copenhagen, in part to move things forwards with the new US team.

Regardless, it could be hard for Obama to make meaningful commitments at the international table without having made significant progress at home. His address last week

came as a surprise; given the ongoing financial crisis, many had expected him to shy away from specifics on climate in his first year in office. During his campaign, Obama largely avoided talking about climate regulation, instead focusing on the more palatable assessment that 'green jobs' in the energy sector could help revive a stalled economy.

Battle on two fronts

Shortly after Obama's speech last week, Boxer announced plans to introduce in January a climate bill directing the Environmental Protection Agency (EPA) to establish a cap-and-trade programme. This leaves many of the details that would otherwise be negotiated in Congress up to the EPA, but the legislation itself would be much simpler and could be ready for early action next year. In the House of Representatives, Democrats also accelerated the path toward climate legislation by replacing John Dingell of Michigan, a long-time ally of Detroit's automobile industry, with liberal environmentalist Henry Waxman of California as chairman of the powerful Committee on Energy and Commerce.

Obama will need to negotiate simultaneously on both the domestic and international fronts, keeping in mind that the Senate must ultimately approve both domestic legislation and any eventual international treaty. He is unlikely to want to negotiate a treaty on his own, the way President Bill Clinton did with Kyoto, and then try to sell it to senators; nor will he want Congress to pass a bill that would effectively limit

his ability to deal on the international stage.

Timothy Wirth, president of the United Nations Foundation, a public charity in Washington DC, also sees potential for the president to directly engage with China and find ways for the world's two largest polluters to work together. "It's like a ladder: on one side of the ladder is the United States and on the other side is the Chinese and you are building these rungs between the two," he says. "Slowly but surely, by the time you get to Copenhagen, you've got the two countries with common interests."

Although China is aggressively pursuing renewable energy and efficiency gains in its heavy industries, it has so far refused to make any firm commitments on reducing greenhouse-gas levels. Pan Jiahua, who heads the Research Centre for Sustainable Development at the Chinese Academy of Social Sciences in Beijing, says he does not see that situation changing before Copenhagen. "China is a very active player on the climate negotiations front," he says. "China is hesitant to make any concrete commitments, but that does not mean it is not active."

Robert Stavins, an environmental economist at Harvard University, says negotiators could sign an agreement on some basic principles in Copenhagen and then work out the details in subsequent years. He adds that it took three years to flesh out the Kyoto protocol after it was signed.

"As in any negotiation, what you want to start with is what you agree on," Stavins says. "If you start with what you disagree on, you don't get anywhere." ■

K. CAMPBELL/GETTY

Acoustic sensors for rare porpoise

SAN FELIPE, MEXICO

After two months of laying down acoustic devices across the Gulf of California in Mexico, three vessels have completed a survey of the world's most endangered marine mammal.

The sophisticated array tracked the estimated 150 remaining porpoises, known as vaquitas (*Phocoena sinus*). The research, completed this week, could help authorities to prevent the last of the vaquita from becoming fishing by-catch. The team eventually hopes to establish a network of buoys around a broad reserve, to serve as a model for porpoise conservation around the world.

Mexican authorities have tried to protect the vaquita by buying up hundreds of fishing skiffs, but the fishermen always return. The vaquita is ranked 78th on the EDGE list of Evolutionarily Distinct and Globally Endangered mammals.

In 1997, a visual survey estimated the vaquita population at nearly 600 porpoises. Each vaquita can be up to 1.5 metres long and weighs about 50 kilograms (A. M. Jaramillo-Legorreta, L. Rojas-Bracho and T. Gerrodette *Mar. Mamm. Sci.* **15,** 957–973; 1999). Over the past three years, scientists have used a cabin cruiser converted into an acoustic lab to follow the animals' sounds — and listened in as the number of vaquita sonar clicks dropped dramatically.

R. DALTON

T. A. JEFFERSON

Marine biologists have been deploying hydrophones (above) to pick up the vocalizations of vaquita (left) in Mexico's Gulf of California.

Sound trackers

These observations provided the impetus for the new survey. Small and understandably nervous of boats, vaquita are difficult to find, especially when the Gulf of California's high winds turn a bathtub surface into a choppy mass of white caps. On a clear day, the best equipment can capture a signal only if the animal is within a couple of hundred metres of the boat. "We have been lucky finding them," says co-chief scientist Lorenzo Rojas Bracho, who coordinates Mexico's Marine Mammal Research and Conservation project from the National Institute of Ecology's office in Ensenada. Even so, he and his co-workers often go for days without seeing any.

Their study is the most sophisticated to date and involved deploying four types of hydrophone. During the surveys, one hydrophone was towed behind the vessels along transect lines, and the others were deployed on buoys or hung overboard for 12–24 hours at a time at multiple locations in the shallow waters of the gulf.

The buoy hydrophones, powered by off-the-shelf batteries, were encased in plastic pipe weighted to the sea floor. Crew members aboard the Mexican cruiser, the *Koi Poi*, retrieved them about every two weeks. From there Rojas Bracho's colleagues, Armando Jaramillo Legorreta and Gustavo Cardenas Hinojosa, downloaded the acoustic records to an onboard computer before resetting and resubmerging the hydrophones. They then sent the data to a nearby vessel, whence they were e-mailed to institutes in Japan, the United Kingdom or La Jolla, California.

If it works as well as its architects anticipate, the array could be the next step in expanding monitoring efforts in other regions. For instance, one of the buoy-deployed acoustic monitors was originally developed by team member Tomonari Akamatsu of the National Research Institute of Fisheries Engineering in Ebidai, Japan, and first used to study fin-less porpoises in Asia. Those monitors have also been laid in the Yangtze River in China to try to find the baiji (*Lipotes vexillifer*), which was declared probably extinct in 2006. And earlier this month, Akamatsu put them in a reservoir behind a dam on the Ganges River in India, to count the endangered Ganges River dolphin (*Platanista gangetica*).

Even with the monitoring, the vaquita's numbers are so low that it may yet go the way of the baiji. Fishermen along the gulf's broad, white beaches insist that there are no vaquita left, in the hope that the government will not restrict their activities. When they are shown videos of the animals in the wild, they call it a computer trick.

First results from the new survey are expected to be presented at a conference in Mexico early in 2009. Then Mexican officials will use it to design a system to monitor, and perhaps prevent, the extinction of the vaquita. ■
Rex Dalton

"We have been lucky finding the vaquita."

Mice share yeast's ageing system

A protein that regulates lifespan in yeast by protecting its genome may perform the same function in mammals, studies in mice suggest. The work addresses a lingering uncertainty about sirtuins, a family of proteins that are a focus of research in the lucrative market for anti-ageing drugs. In April, Sirtris Pharmaceuticals, based in Cambridge, Massachusetts, and co-founded by David Sinclair of Harvard Medical School, who is the lead author of the new work, was purchased by pharmaceutical giant GlaxoSmithKline for US$720 million.

Researchers already knew that a yeast sirtuin called Sir2 is a sentinel of the organism's genome, preventing genes from being expressed at the wrong times and blocking the chromosomal rearrangements that sometimes occur in areas of repetitive DNA sequence. But when DNA strands break, Sir2 molecules move to repair the damage, leaving their usual positions unguarded. As cells age, the rate of DNA damage increases, forcing Sir2 proteins to leave their original posts more frequently. Some genes that were meant to be silenced are then free to be expressed, generating a shift in patterns of gene expression that is characteristic of ageing.

Increasing Sir2 levels slows ageing in yeast, but it has been unclear whether sirtuins would act by a similar mechanism in mammals. "People, including myself, struggled with the acceptance that a mechanism of ageing in a yeast could be relevant to a human," says Sinclair.

Now, he and his colleagues report that a mouse sirtuin called SIRT1 behaves much like its yeast counterpart (P. Oberdoerffer *et al. Cell* doi:10.1016/j.cell.2008.10.025; 2008). In mouse embryonic stem cells, SIRT1 also associates with regions of repetitive DNA, and silences the expression of certain genes. But when the cells are treated with hydrogen peroxide, a chemical that can cause DNA damage, SIRT1 is recruited to the site of DNA breakage, and previously silenced genes become expressed. Furthermore, the team found similar changes in gene-expression patterns in brains of elderly mice. "This could be the key to understanding how sirtuins could be beneficial during ageing," says Jan Vijg of the Albert Einstein College of Medicine in New York.

Sinclair now aims to find out whether increasing SIRT1 expression in old mice can revert gene expression to a more youthful pattern — effectively reversing the effects of ageing. ∎
Heidi Ledford

David Sinclair wants to reverse the effects of ageing.

R. FRIEDMAN/CORBIS

Social-science programme hit by murder charge

A civilian employee of the US Army's highly controversial 'Human Terrain' social-science programme has been charged with second-degree murder.

Although originally proposed as a means of improving relations between the military and locals in Afghanistan and Iraq, the Human Terrain programme has become mired in arguments about how much academics should be assisting military operations (see *Nature* 455, 583–585; 2008).

According to an affidavit filed in a Virginia court, American Don Michael Ayala shot Afghani civilian Abdul Salam in the head while on patrol with the US Army in Afghanistan. At the time, Salam was restrained and handcuffed after attacking and badly injuring Paula Loyd, another member of the Human Terrain Team. Two Human Terrain researchers have died this year: Michael Bhatia in Afghanistan in May, and Nicole Suveges in Iraq in June.

Distant world sports gassy signature of habitability

Carbon dioxide has been spotted in the atmosphere of a gas-giant planet called HD 189733b, which orbits a star 63 light years from Earth.

Along with water and methane, which have already been detected in the atmospheres of other extrasolar planets, the presence of carbon dioxide could give astronomers clues about whether a planet might support life.

Mark Swain of NASA's Jet Propulsion Laboratory in Pasadena, California, reported his team's findings on 19 November at the Paris conference on Molecules in the Atmospheres of Extrasolar Planets.

The observations were made using the NICMOS near-infrared camera on the

C. CARREAU/ESA
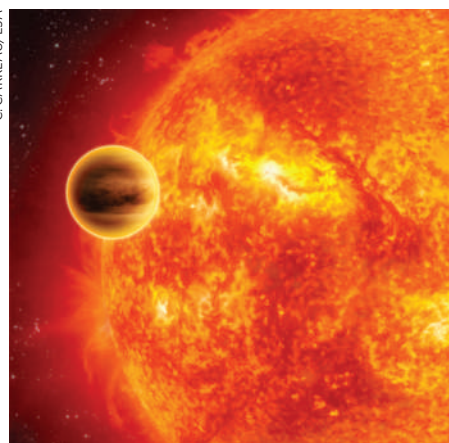
An artist's impression of planet HD 189733b.

Deep-ocean observatories take the plunge

A deep-water ocean observatory began operations off the west coast of North America earlier this month, making it the second to go live in the area this year.

On 10 November, the Monterey Bay Aquarium Research Institute in California began receiving data from its US$13.5-million Monterey Accelerated Research System (part of it is pictured being lowered into the sea) — a platform the size of two small cars to which researchers can link their recording instruments to monitor earthquakes, sea acidification and animal life at a depth of 900 metres.

K. FULTON-BENNETT/MBARI

Meanwhile, Neptune Canada, which aims to be the world's largest cabled seafloor observatory, saw data from its first instrument node on 29 September. By the end of 2009 it should boast five nodes along an 800-kilometre cable loop, reaching depths of 2.7 kilometres to measure, for example, current speeds, water pressure and sonar signals.
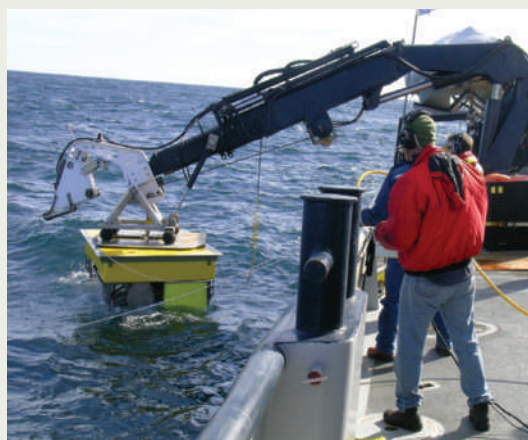
Hubble Space Telescope. The detection of carbon dioxide is particularly surprising, because other forms of carbon, such as carbon monoxide and methane, would be expected to dominate the planet's hot, hydrogen-rich atmosphere.

European Union declares Arctic interests

The member states of the European Union (EU) have joined those countries staking an official interest in the Arctic, after the EU last week released a declaration of its "interests and policy objectives" in the far north. The European Commission billed it as the "first step towards an EU Arctic Policy".

The move comes amid steadily rising rhetoric from nations bordering the Arctic, with Russia, Canada and the United States rattling sabres over their rights to the area. As well as support for climate-change mitigation and oil exploration, the declaration also renews the EU's commitment to building a new research icebreaker, *Aurora Borealis*.

Further projects are promised that will monitor rising sea levels and the loss of sea ice and permafrost, as well as screening for organic chemicals and heavy metals in the region.

United Kingdom auctions carbon emissions permits

The United Kingdom raised nearly €65 million (US$82 million) last week in the first auction of carbon-emission allowances in the second phase of the European trading scheme.

Industries that are part of the scheme need one carbon allowance for every tonne of carbon dioxide they emit into the atmosphere. These allowances have been distributed for free until now, but European Union (EU) rules allow governments to auction up to 10% of the permits between now and 2012, when phase two ends.

The country sold 4 million allowances at a rate of €16.15 per tonne on 19 November, marking a transition towards a system in which companies will be required to pay for the right to pollute. Other EU countries are preparing to follow suit, and the EU is debating rules that would require all carbon allowances to be auctioned from 2013.

Unique collaborations at Howard Hughes institute

The Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, has launched a pilot programme that will see HHMI money flowing directly to non-institute scientists for the first time.

Through four-year Collaborative Innovation Awards, the HHMI will spend US$40 million to fund eight teams of three to six scientists, headed by current HHMI investigators, which will tackle novel biological problems.

One of the projects aims to combat glaucoma, a leading cause of blindness, by developing a miniature, implantable wireless sensor that continuously monitors pressure within the eye. Another will try to identify a drug that can clear out the brain proteins thought to cause neurodegenerative disorders such as Alzheimer's disease. The institute expects to expand the programme in coming years.
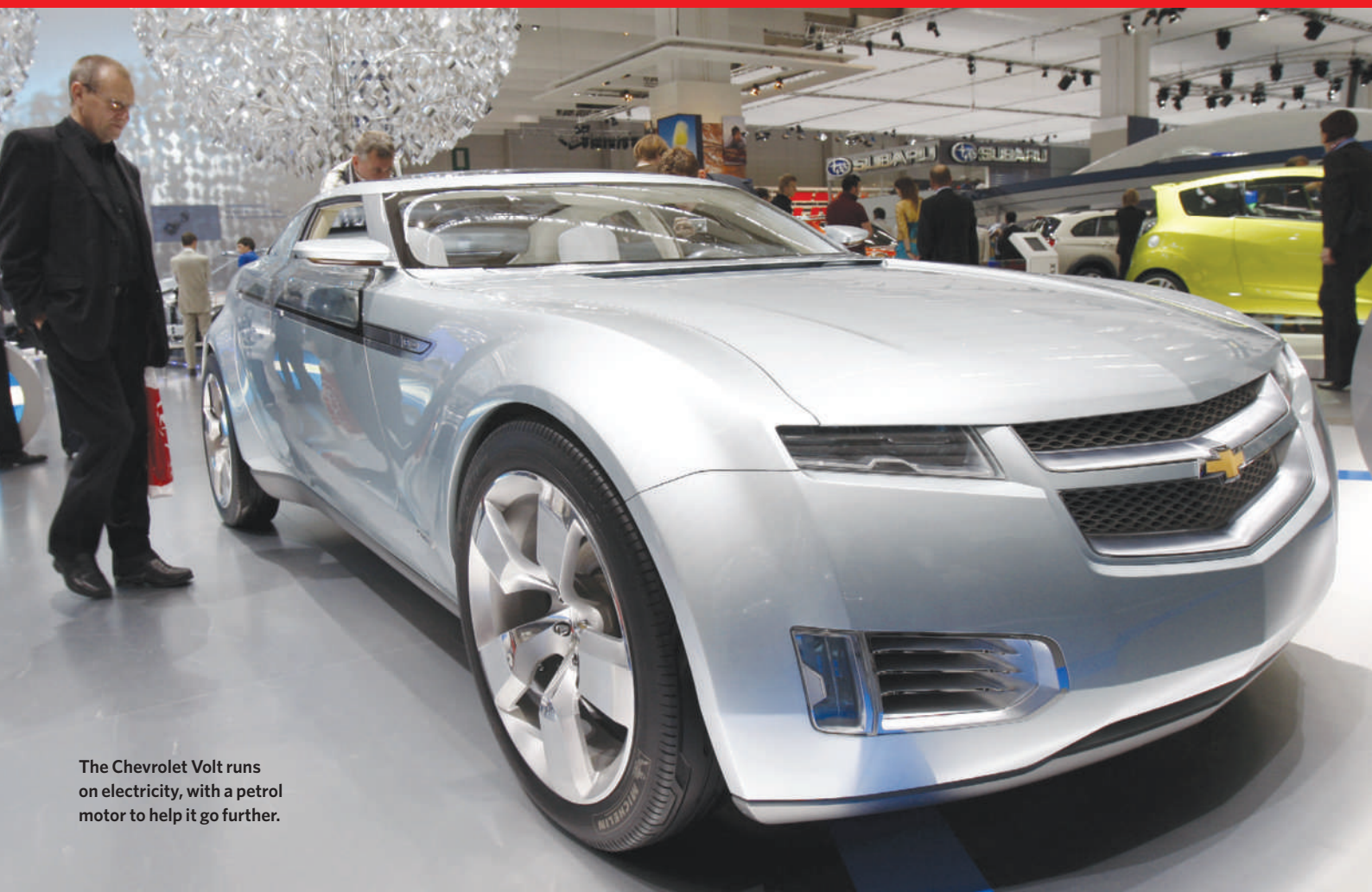
The Chevrolet Volt runs on electricity, with a petrol motor to help it go further.

CHARGING UP THE FUTURE

A new generation of lithium-ion batteries, coupled with rising oil prices and the need to address climate change, has sparked a global race to electrify transportation. **Jeff Tollefson** investigates.

"We have had a massive shift in one of the biggest industries in the world," says Stephan Dolezalek, who leads the CleanTech group at the venture-capital firm VantagePoint Venture Partners in San Bruno, California. Dolezalek has been watching the global automobile sector embrace the idea of plug-in electric cars: "In three years we've gone from thinking 'it can't be done' to not only 'it can be done' but 'we are all going to do it.'"

The shift is partly a story of technological innovation, which has produced rechargeable batteries that pack enough power to propel some of the basic passenger vehicles currently being designed further than 200 kilometres. Billions of dollars have poured into start-up companies that promise new batteries, and billions more have poured into fledgling electric-car manufacturers eager to take on the global automotive giants — every one of which is also developing electric vehicles.

The shift is also a story of oil supplies, national security and global warming. Record-high oil prices have pushed consumers towards fuel-efficient vehicles and prompted many governments to consider electric transport as a way to escape their dependence on imported petroleum and to address climate change. Money currently spent abroad could instead be spent on domestic power generation from wind, solar and other low-carbon energy sources.

And the shift is a story of a shared vision: developing the technology that would entice all drivers to plug in rather than fill up. Millions of battery-powered cars plugged into an increasingly green electric grid would not only save drivers money and reduce greenhouse-gas emissions, it would also provide the grid with a distributed, high-capacity storage system for electricity. Such a system would help to accommodate the variable and unpredictable nature of renewable electricity sources. And further out, it could allow power companies to store energy generated during times of low demand, then draw it back again to meet peak demand. The end result could be more a stable and efficient grid that might even lower home electricity bills.

Getting there won't be easy. All these hopes hinge on battery technology that is only just emerging from the lab. A suite of technical challenges remains to be overcome, and it is not yet clear how much further the technology can be pushed. At the same time, the manufacturers who are arguably best able to bring about these changes — the global automotive giants — have been hammered by an energy crisis followed by an epic financial meltdown.

None of them has abandoned the effort yet, in large part because they all believe that, despite the current lull, oil prices have nowhere to go but up. Moreover, batteries have leapt ahead of expensive hydrogen fuel cells as the technology of choice for getting beyond oil, at least for now. But the field is wide open in terms

of bringing them to market. Dolezalek believes that major car companies might well perish in the face of versatile young upstarts, and he isn't alone. The automobile industry secured a place in this autumn's first round of economic bailouts from the US government with US$25 billion in loan guarantees for retooling its plants, and it is already seeking more. That has people such as Andrew Grove, former chairman of Intel, who has become a leading proponent of electric transportation, talking about the 'valley of death' that often accompanies a massive technological transformation. Grove says that car manufacturers have already begun their march through the valley, knowing that many won't make it through to the other side.

"The only time people make these moves [through the valley] is when things are rough, but they can't afford to make them when times are rough," Grove says. And that means that governments might have to step in. "I just hope that it's going to be done in such a way that the government says, 'I'll give you some water and food to get through the valley of death, but don't turn back.'"

Building a better battery

Pioneers have turned back before, most notably General Motors. In 1996, the US company released the EV-1, the first all-electric car from a major manufacturer. The vehicle was expensive, rolled out in response to a California mandate, which was later rescinded, for 2% of all cars sold in the state to have zero-emissions by 1998. But its fate was ultimately sealed by one thing: its battery.

Building batteries has been an exercise in chemical compromise for more than two centuries. The idea is simple: chemical bonds can be used to trap ions in one electrode. When a battery is hooked up to a circuit, the ions flow through a separator to a second electrode; as the ions flow, they release electrons, generating an electric current. In rechargeable batteries, the chemical reaction can be reversed to store energy (see graphic, overleaf). But the reality is complex: although scientists have produced numerous potential battery chemistries (see *Nature* **451,** 652–657; 2008), none of them performs well on all the crucial factors of cost, safety, durability, power and sheer capacity.

The first-generation EV-1 deployed a lead-acid battery, still the technology of choice for conventional vehicles. Lead-acid batteries are safe, cheap, long-lived and reliable, but they are also big and heavy. They could push the car for about 150 kilometres per charge. A second-generation vehicle released in 1999 featured a nickel metal hydride battery, and travelled 50% farther on a charge, but General Motors

> "Don't worry about charging electric cars from some perfect grid of the future — just get the cars out there."
> — Mark Duvall
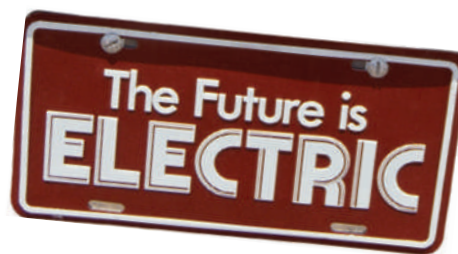
cancelled them after the first year, saying that it could not sell enough to make them profitable.

It was a decision that General Motors would come to regret. As it turned back to large and profitable vehicles such as the Hummer, its up-and-coming Japanese rival Toyota was digging into the new technology, using the same battery that General Motors had abandoned to produce hybrid cars that combined a standard combustion engine with an electric motor. Toyota has gone on to set the standard for hybrids: its third-generation Prius has been immensely popular, proving that consumers will adopt advanced battery technology in automobiles if it is done well. The Prius fortified Toyota's reputation, and helped it to surpass General Motors last year to become the largest automobile manufacturer in the world.

But nickel metal hydride batteries can be developed only so far. These batteries pack more power than standard lead-acid ones but can be permanently damaged if allowed to discharge too far. To maintain an adequate safety margin, Toyota limited the Prius to using about 20% of its battery charge during normal operation. But although not using 80% of the capacity is acceptable if the battery is simply supplementing a petrol engine, it is a luxury that fully electric cars can't afford. Electric cars need all the charge they can get, and that means new chemistries.

Lithium-ion batteries, which are compact and have a high capacity, are a natural place to start. Sony paved the way with the lithium cobalt oxide battery, which made its mass-market debut in a 1991 version of the firm's Handy-Cam video camera, and is now widely used in consumer electronics. Lithium is a light metal, and the lithium cobalt oxide lattice structure allows plenty of space for the give and take of ions. But scaling this chemistry up for vehicles is problematic. Cobalt is expensive and toxic, and the batteries have been known to show 'thermal runaway', battery lingo for fires or explosions. "It has affected a tiny, tiny fraction of all of the batteries sold, but nonetheless, it's pretty freaky to think about a big fire in one of the vehicles," says Jeff Dahn, who works on advanced battery technology at Dalhousie University in Halifax, Canada. "Safety really needs to be the focus for the research community."

Many of the lithium batteries under development for vehicles replace cobalt oxides with manganese oxides and iron phosphates. Both are safer, but they do have their own problems, not least of which is a lower storage capacity for their size. Another challenge has been dealing
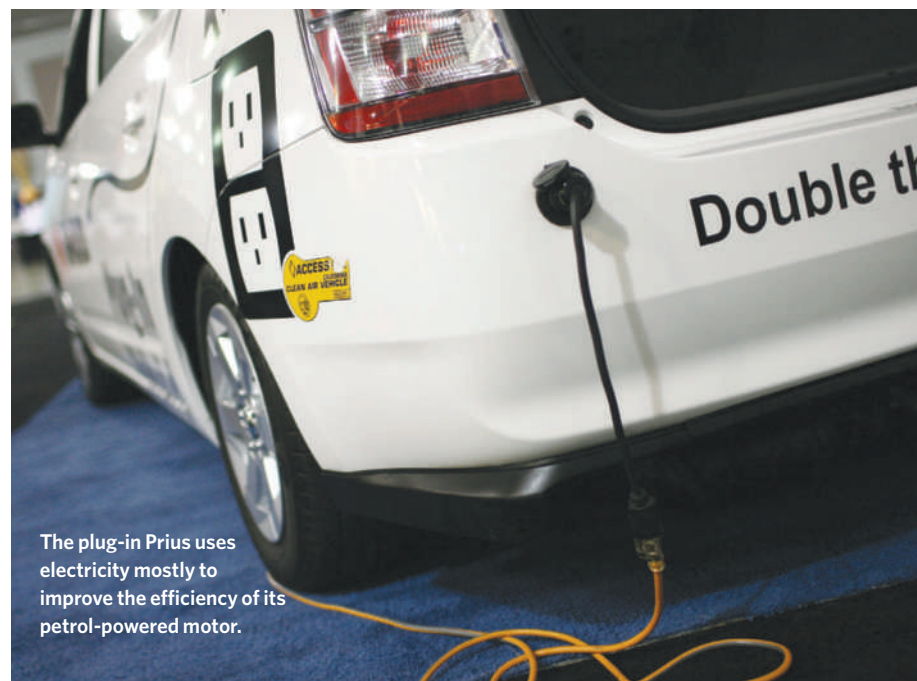
The plug-in Prius uses electricity mostly to improve the efficiency of its petrol-powered motor.

with the physical expansion and contraction of the electrode material as the lithium ions flow back and forth during charge and discharge, which can lead to fractures. Researchers at multiple institutions have addressed the issue by adding carbon and other substances to the electrode material.

They are also probing other chemistries — often at the nanoscale — based on silicon, fluorides and oxygen, which have a greater capacity. Others are looking at equipping the battery pack with capacitors, which can rapidly store and discharge electricity.

Even in their current state, however, lithium-ion batteries are performing well enough to keep car manufacturers interested. Last year, General Motors inaugurated the race for mass-market electric vehicles when it announced plans to market its plug-in hybrid, the Chevrolet Volt.

A break with the past
The Volt, now scheduled for a 2010 roll-out, is a radical shift in design. Hybrids such as the Prius are powered by petrol, and use a battery simply to improve fuel efficiency. The Volt hybrid will be the reverse: an electric car that uses petrol to extend its range. Only when the charge dies will a small petrol motor kick in to charge the battery, which then continues to power the vehicle. The goal is for Volt owners to plug in at night and then drive more than 60 kilometres a day on a single charge — before burning a single drop of petrol. Given that as many as 80% of US drivers commute less than that on an average day, such vehicles could eliminate a sizeable chunk of the nation's oil consumption.

The Volt initiative could open the door to a new kind of transportation system — if the company can pull it off, both on time and at a cost that will tempt consumers. Many observers have their doubts. "They are fundamentally redefining what a car is, but can they do it? I don't know," says Don Hillebrand, who heads the Center for Transportation Research at Argonne National Laboratory in Illinois. "When the first generation of anything comes out, to a certain extent car manufacturers are rolling the dice, and this is the biggest roll of the dice anybody has ever made."

Some say it is a long shot. With sales plummeting in the midst of a deepening recession, the company is facing possible bankruptcy, and has joined with the other major US car manufacturers in seeking an additional bailout from the government. But through it all, General Motors has continued to sink everything it can spare into the Volt, viewing it as a

key technology that would allow the company to leapfrog its competitors.

Toyota is taking a more measured approach with its plug-in hybrid, which is expected to roll out with a lithium-ion battery in 2009. John Hanson, a spokesman based at Toyota's US headquarters in Torrance, California, talks about managing customer expectations: the company is promising only that the vehicle will go "at least" 16 kilometres on an electric charge. After that, it will blend petrol and electric power in much the same way as the current Prius.

That would leave General Motors in pole position, at least in terms of the electric range it is promising. But will the Volt succeed? The answer to that question depends on consumers. What will they want several years from now? And how much will they be willing to pay? General Motors expects to lose money in the beginning and has not yet announced a price for

> **"Safety really needs to be the focus for the research community."**
> — Jeff Dahn

LITHIUM-ION BATTERY

Charge mechanism

Negative electrode (carbon)

Electrons

Positive electrode (lithium compound)

Li⁺

Separator

Cell filled with electrolyte

Discharge mechanism

Electrons

Negative electrode (carbon)

Positive electrode (lithium compound)

the vehicle, but the continued viability of the firm could depend on how fast it can sell the new cars and at what price. The company is banking on tax credits, enacted this year by Congress, to encourage people to buy plug-in hybrids, and high petrol prices would help as well.

But the firm's chief economist Mustafa Mohatarem says that he can't help but wonder whether consumer demand for electric vehicles has been exaggerated. "It is critically dependent on the battery," he says. "Until you have a much better handle on the cost of this technology, to talk about demand is in a sense ridiculous."

Have we forgotten something?
Others look at the market and see a different problem: a lack of batteries. Charles Gassenheimer, chief executive of Ener1 Group, a company in New York that produces lithium batteries, says that car manufacturers have collectively announced some 75 types of electric cars that are supposed to hit the road by 2013. But they have been slow to commit to orders, he says. And without orders, battery manufacturers can't invest the time and money necessary to ramp up production, a bottleneck that could delay the roll-out of new vehicles.

Governments seeking to spur the electric-car market must look at battery manufacturing in addition to consumers and car manufacturers, Gassenheimer says. "There needs to be some government intervention at this phase in the game. Otherwise it's going to be a chicken-and-egg problem that doesn't get solved."

Gassenheimer also raises concerns that countries such as the United States will simply trade their dependence on Middle Eastern oil for a reliance on Asian batteries. He has a sizeable stake in the outcome, of course, but the issue has political resonance as governments look to spur new green jobs. Experts say that Ener1 and other Western companies have the technology, but Asian companies have a leg-up on the manufacturing side simply because Asia has such a lead in producing lithium-ion batteries for electronics.

"The United States is certainly not being blindsided at this time, but whether or not we really have the resources and critical mass to compete in the long term in automotive batteries is still very much an open question," says Yet-Ming Chiang, a materials scientist at the Massachusetts Institute of Technology in Cambridge, and founder of lithium-battery manufacturer A123Systems in Watertown, Massachusetts. "The same thing goes for Europe."

Others dismiss concerns about where the batteries are going to be made, citing a crucial difference between electronics and vehicles:

New-generation lithium-ion batteries can power a car for up to 200 kilometres.
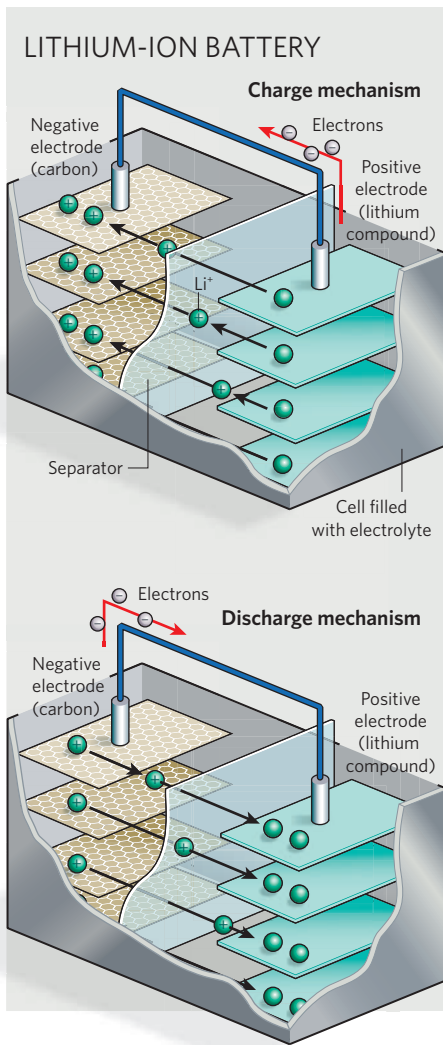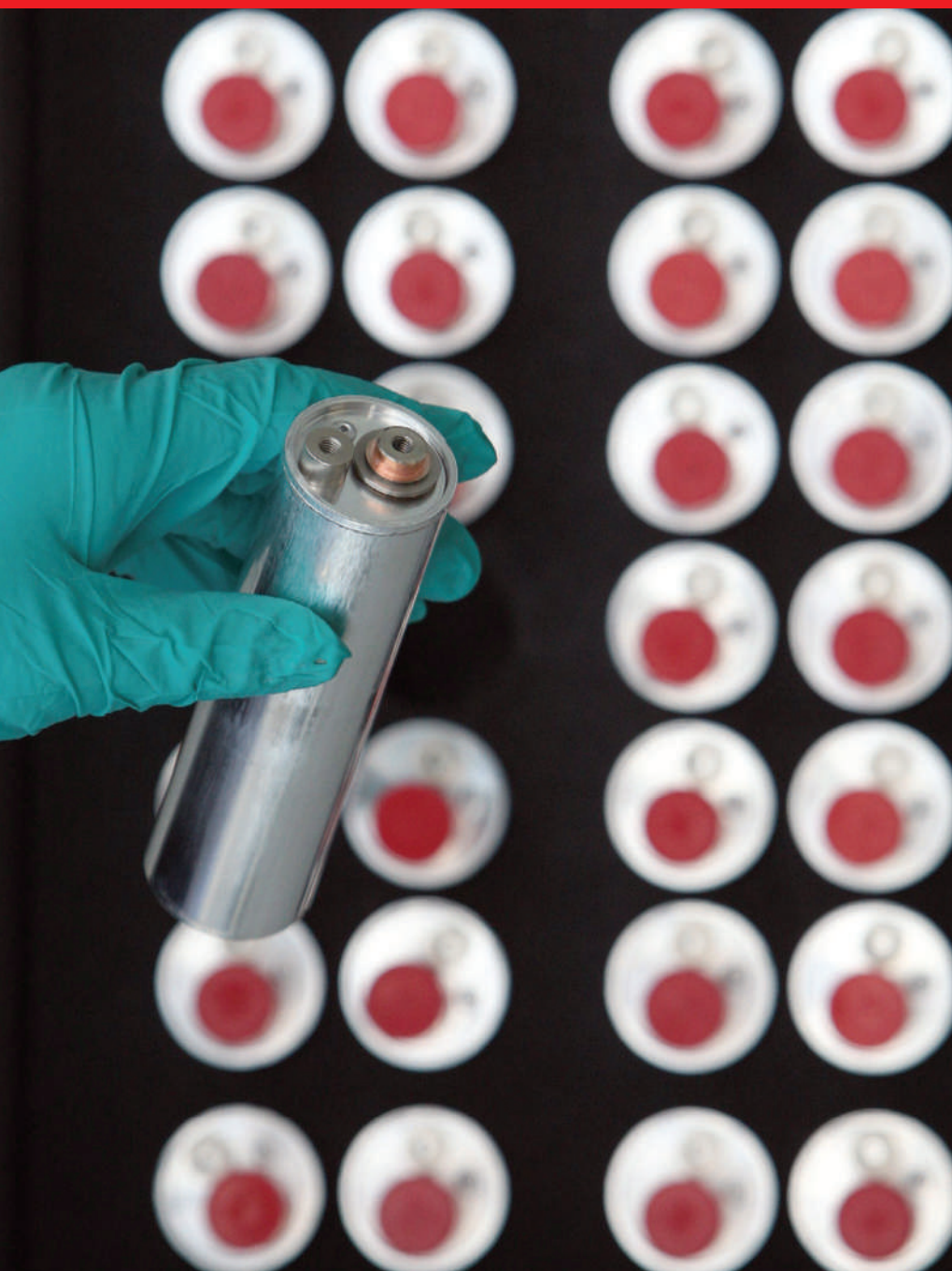
Making connections

So what will the market be like when it does come? Plug-in hybrids such as the Volt represent a leap beyond battery-augmented cars that merely make better use of petrol. They also give drivers the freedom to run on electricity for short trips while still making long trips, albeit guzzling gas on the way. But some car manufacturers say that the best path forward would be an all-electric vehicle, which could one day all but eliminate oil consumption in the transportation sector.

Getting rid of the petrol motor greatly lessens costs and complexity and opens up space for more battery power. "In terms of a solution, both from a carbon dioxide point of view and from a technical point of view, the hybrid and the plug-in hybrid do not provide the technical breakthrough that the electric vehicle could provide," says Serge Yoccoz, who is in charge of electric vehicles at Renault. "And from what we've seen, the plug-in hybrid is definitely more expensive [than an electric car would be], even if you take into account the need to develop a charging infrastructure."

So while researchers search for the technical breakthrough, entrepreneurs are trying to get around the high costs by rethinking the way we market cars, batteries and ultimately energy.

One such innovator, Better Place of Palo Alto, California, is aiming for nothing short of a wholesale conversion of the transportation sector. The company likens itself to a cellphone network for all-electric cars: you buy the car from a Better Place partner and then sign up for one of its various user plans. Better Place then provides a network of charging spots — at home, work and retail outlets — as well as stations at which used battery packs could be swapped for recharged ones by a robotic arm in a matter of minutes (see graphic, overleaf).

But to accomplish all this, Better Place needs a computer system that can track electricity charges wherever they are incurred. It also needs to partner with governments and industry, including the automotive, battery and utility sectors. So far, Better Place has lined up partnerships with an alliance between Nissan and Renault to pursue electric cars, and the company plans to roll out its system in Israel, Denmark, Australia and California, with the first deployments scheduled for 2010.

The scheme is ambitious, but Sidney Goodman, head of automotive alliances at Better Place, says that's the only way to do it. "We don't believe we can do this on a small scale. It's one of these projects where either you do it big or you don't do it".

Better Place is aiming to provide family sedans that have a 160-kilometre range in an effort to attract all drivers, not just city commuters with

electronics are by and large made in Asia, but cars are made in the West, too. Building batteries near automobile plants would not only save money, it would also get around complex international shipping regulations that put lithium-ion batteries in the 'dangerous goods' category. The market "is driven by where the end product is", says Khalil Amine, a battery researcher and one of Hillebrand's colleagues at Argonne. "For electronics, we buy everything from Asia. For transportation, there is plenty of production here."

General Motors tested lithium batteries from every manufacturer it could find and narrowed the decision down to two companies: A123Systems and LG Chem, a Korean giant that made its name in electronics. Only in late October did the contract reportedly go to LG Chem,
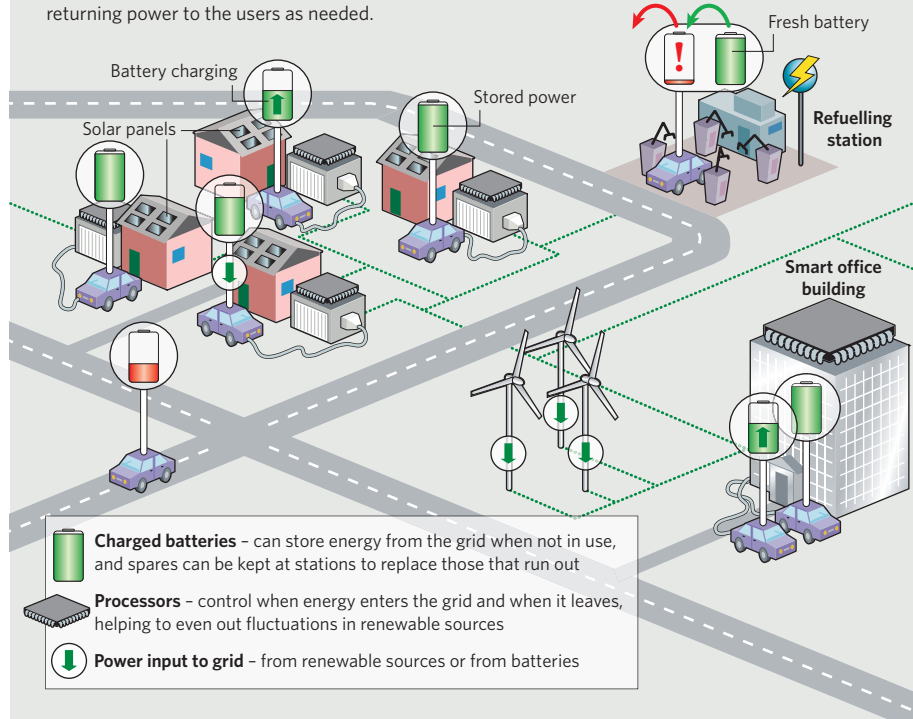
which has a stronger base and a longer history on the manufacturing side. LG Chem has already partnered with the Korean car manufacturer Hyundai to supply 7,000–10,000 lithium-battery packs for a pair of hybrid vehicles that will begin rolling off the line in 2009.

Soonho Ahn, LG Chem's vice-president for battery research and development, says that his company isn't expecting to make money on its automotive batteries for some time but wants to be ready when the market takes off. He notes that the battery market in Asia is in "equilibrium" after several years of stiff competition in the electronics sector. "We have some time to look at the next mega-application, and the next mega-application is the automotive industry," he says. "We're pretty sure that the market is coming."

THE INTEGRATED GRID

Electric cars could provide a distributed storage system for the grid, returning power to the users as needed.

Battery charging

Solar panels

Stored power

Fresh battery

Refuelling station

Smart office building

Charged batteries – can store energy from the grid when not in use, and spares can be kept at stations to replace those that run out

Processors – control when energy enters the grid and when it leaves, helping to even out fluctuations in renewable sources

Power input to grid – from renewable sources or from batteries

an environmental bent. Goodman runs through some rough numbers — assuming that a battery costs US$15,000 (which is likely to be on the high end of the scale, he stresses), an electric vehicle would cost about 6 cents per kilometre to power. That compares with just under 12 cents per kilometre for conventional cars in the United States, and twice that in Europe.

A Norwegian company called Th!nk is taking a similar route with its all-electric commuter car, which is due to hit roads in Norway, Denmark and Sweden in coming months. With an initial price tag of about 200,000 Norwegian krone ($30,000), the car will cost about 20% more than the same-sized petrol-powered car and will drive some 180 kilometres on a charge. Customers then pay a monthly lease to cover the cost of electricity and the battery. "We'll get the costs of our car down to somewhat similar to the cost of a petrol-powered car, and then we'll have a very strong proposition going forwards," says Richard Canny, Th!nk's chief executive.

Tapping the matrix

Utility firms are eager to cooperate. Although making electric vehicles a reality will require unprecedented cooperation between two industries that have until now had little in common, utilities actually see many more benefits than headaches. The fundamental

fact is that most of the charging would take place at night, which creates a new source of revenue at a time when utilities typically have excess capacity.

In the end, this should translate into substantial reductions in greenhouse-gas emissions, even in countries such as the United States that get much of their electricity from coal. A plug-in hybrid running on electricity generated entirely from coal is roughly equivalent to a conventional hybrid in terms of emissions, but utilities say that in the early years, electric vehicles will frequently draw power from spare generating capacity that uses cleaner-burning natural gas. Scaled up, millions of batteries — either in cars or in a future after-market for used batteries — could provide utilities with a flexible storage system that could soak up renewable power, particularly from wind turbines at night.

Assuming that plug-in hybrids will make up 60% of the US automobile market in 2050, electric transport would consume as little as 8% of the nation's electricity, according to a joint modelling study conducted in 2007 by the Electric Power Research Institute — a non-profit research organization in Palo Alto —

and the Natural Resources Defense Council, an environmental advocacy group based in New York. The resulting report, *The Power to Reduce CO_2 Emissions*, predicts that the nation would use 15–20% less oil and reduce its greenhouse-gas emissions by 450 million tonnes, which is akin to pulling 82.5 million internal-combustion vehicles off the road.

"Our fundamental conclusion from this study is that the number one driver of benefits is really the number of vehicles," says Mark Duvall, programme manager of electric transportation at the Electric Power Research Institute. "Don't worry about charging them from some perfect grid of the future — just get the cars out there. They don't have to be perfect."

Utilities such as Southern California Edison in Rosemead are already thinking about how to integrate cars into the electricity system, allowing them to charge up at work or park in 'smart garages' that coordinate activities between the car and the grid. In the early days, advanced charging equipment would communicate with the utility to time the charging so that everybody's vehicle is fully juiced when it needs to be — but not necessarily before. That would help ensure that millions of vehicles don't create a sudden surge on the electricity system when people return from work, when they also tend to turn on lights and crank up their appliances. Further out, this process could be reversed, allowing batteries to provide power to the grid when it is needed most — so long as they are fully charged when it comes to time to drive.

Levelling out the daily demand cycles would allow utilities to manage the grid more efficiently, potentially lowering costs to consumers. "The more cars that come onto the energy system, the better off it is for the energy system," says Ed Kjaer, director of electric transportation for Southern California Edison. And Kjaer says that the vehicles will become cleaner over time as utilities expand their renewable electricity offerings.

> **"We have some time to look at the next mega-application, and the next mega-application is the automotive industry."**
> — Soonho Ahn

That kind of logic has convinced many researchers that electric cars are a must if the planet is to deal with global warming, even if they ultimately raise the stakes on efforts to produce carbon-free electricity. "We've got to electrify the transportation system and then clean up the grid," says Timothy Lipman, research director at the University of California's Transportation Sustainability Research Center in Berkeley. "It's the easiest path." ■

Jeff Tollefson covers climate, energy and the environment for *Nature*.
See Editorial, page 421.

Fresh faces: Cristina Fernández de Kirchner (left) and Lino Barañao plan major scientific investment.

THE COME BACK

Argentina's government has pledged to reverse a decades-long scientific brain drain. **Rex Dalton** reports.

In September 2007, with a wind of change already in the air, about 40 people arrived at Argentina's New York consulate for an unusual meeting. Nearly all of them were Argentinian expatriate scientists, now conducting research abroad. The organizer, mathematician Adrián Paenza, was a famous football journalist turned Buenos Aires talk-show host. The guest of honour was Cristina Fernández de Kirchner, the woman almost assured of winning Argentina's presidential election the following month.

For two and a half hours, Kirchner held court. She listened as the researchers suggested how to improve science in their homeland, and she peppered them with questions about black holes, applied mathematics and computer science. She had really done her homework, several of the attendees noted. Kirchner wanted to know what would be needed to entice researchers back to Argentina — and to recapture a level of scientific excellence that some say was last seen in 1947 when Bernardo Houssay became Argentina's first scientific Nobel laureate when he won the prize for physiology and medicine.

From the 1960s until the end of the 1980s, Argentina's ruling parties viewed academia as a breeding ground of political dissent. Students and professors were targets for suppression and were among the tens of thousands who 'disappeared'. In the 1990s, when the country began to confront its ugly history, President Carlos Menem had no interest in research and one of his economic ministers, Domingo Cavallo, famously told scientists to go and "wash the dishes". During the Argentinian financial crisis that peaked in 2002, any scientists who had not already fled were hit by a peso devaluation that made it near impossible even to attend an international conference.

Kirchner wants change. She says that she sees science as key to the nation's economic future, and the New York meeting helped to crystallize her ideas. After winning the election in October 2007, Kirchner set up a Ministry of Science, Technology and Productive Innovation and named Lino Barañao, a cell biologist and formerly head of the nation's main research granting agency, as the country's first science minister[1]. The move delighted researchers. Alberto Kornblihtt, a biochemist at the University of Buenos Aires, recalls the stunned response from his family: "The president has appointed a biologist to her cabinet? Maybe Argentina is changing."

Barañao and Kirchner hope to grow the country's investment in science and technology from 0.66% of gross domestic product (GDP) in 2007 to 1% in 2010. This would rival the 0.9% of GDP currently invested by Brazil, one of the biggest-spending Latin American countries. Barañao plans both to reverse the brain drain and to retain scientists. He has promised that in 2009 the government will increase the budget for competitive research grants by 40% over the current year, to around 700 million pesos (US$210 million). More is being pumped into professorial labs and salaries. In 2008, the national research council CONICET increased researchers' salaries by 30% to about $1,000 a month; a far cry from the $450 or so that a full professor made in 1990. Infrastructure expansion, with an initial budget of $50 million, is to start next year.

Laying the foundations

In one of its most ambitious plans, the government is building a $50-million science complex called Polo Científico in the trendy Palermo area of Buenos Aires to house several new research institutes. One will be an international interdisciplinary research centre, focusing on biomedicine, information technology, nanotechnology and the social sciences. Another, a partnership with the Max Planck Society in Germany, will have 200 scientific staff working on molecular biology, neuroscience and cancer when it is completed in 2011. And at the University of Buenos Aires, an $8-million science building is being planned to house the centre where Kornblihtt works: the Institute for Physiology, Molecular Biology and Neurosciences, directed by Osvaldo Uchitel.

But it may take more than money to bring scientists back. In the worst years of military rule, the 1976–83 Dirty War, an estimated 30,000 people disappeared. Subsequent investigations revealed that they were often dumped from military aircraft into the ocean. "We all know someone who disappeared," Kornblihtt says. In the university building containing Kornblihtt's lab, there are floor-to-ceiling banners with names

> "We all know someone who disappeared."
> — Alberto Kornblihtt

and photographs of the disappeared students — a daily reminder of the past.

Kornblihtt was one of those who remained in Argentina through much of the troubled times, aided at times by philanthropic grants. In 2000, he was selected by the Antorchas Foundation of Buenos Aires to receive three consecutive three-year grants worth a total of $900,000. "This changed the quality of my science," he says. The money has allowed him to equip his lab, nurture students and collaborate internationally in his work on alternative splicing[2], the process by which messenger RNA is cut into different forms before coding for proteins.

In 2002, and again in 2007, Kornblihtt received a five-year international fellowship of around $100,000 a year from the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland. The HHMI expanded its international programme to Latin America in 1997 and Argentina now has 14 HHMI-funded researchers. Only Canada receives more of the institute's international awards. "We don't make grants because a nation has difficult economic times," says Jill Conley, director of the HHMI's international programme. "We hold the bar high. But it is gratifying to see that this offers some folks a lifeline."

Anonymous donors

As Argentina's economy has strengthened, philanthropy from national donors has also grown. By 2006, the Leloir Institute Foundation in Buenos Aires had amassed a $25-million endowment from private donations, of which about $3 million is now being used to build a new wing that will house core equipment for the institute's 25 principal investigators. The wing is set to open in 2009, when a symposium is planned around the institute's primary fields of neuroscience, cell biology, cancer and infectious disease. But the main donor wishes to remain anonymous. Maintaining such a low profile is common in Latin American countries, because the wealthy fear undesirable scrutiny or kidnapping.

Located in central Buenos Aires, the Leloir Institute looks like a research centre in any major world capital, from its airy labs to the modern art on the lawn. The foundation was named after Luis Federico Leloir, who won Argentina's second scientific Nobel — for chemistry — in 1970. Institute director Fernando Goldbaum, one of five HHMI fellows there, studies brucellosis, a bacterial infection that has re-emerged in recent years in both humans and livestock. As two of the world's largest beef exporters, Argentina and

The state-of-the-art Leloir Institute in Buenos Aires (left) is directed by Fernando Goldbaum (right).

its neighbour Brazil together lose an estimated $100 million annually to the disease. Last year, working with Roberto Bogomolni of the University of California, Santa Cruz, Goldbaum identified light-sensitive proteins that increase the virulence of brucellosis bacteria[3].

Goldbaum says that high-quality research of this type is already commonplace in Argentina, but often receives little international notice. Last year, Leloir scientists felt that Argentina was improperly lumped together with developing countries[4] in an issue of *Nature Medicine* about 'shoestring' science, which included an article about bamboo microscopes in India[5].

Neuroscientist Alejandro Schinder almost jumps from his seat at the mention of bamboo microscopes. Schinder did postdoctoral research with neuroscientist Fred Gage at the Salk Institute for Biological Sciences in La Jolla, California, and then returned in 2002 to establish his own laboratory at the Leloir Institute. In a recent paper, Schinder, Gage and their team showed that newly born neurons in the adult brain make connections with existing cells, resolving a controversy about whether such cells contribute to working neuronal circuits[6]. As part of this, Schinder's lab used a light-activated ion channel to show how the connections functioned. "You don't publish research like this with bamboo microscopes," says Schinder. "Not a single day do I regret my decision [to move back]. The impact of my science is much greater in Argentina than in the United States. We started a virgin field here."

Goldbaum says that six new principal-investigator positions will be added at the institute in the next few years, funded by the endowment. "We have a strong policy to reverse the brain drain," he says.

> "You don't publish research like this with bamboo microscopes."
> — Alejandro Schinder

Barañao estimates that about 850 of the 5,000 researchers in Argentina are of the world-class calibre that he particularly wants to retain and, to do so, the science ministry plans to match the money that Argentinian researchers receive from international grant givers, such as the HHMI. "We want to reward the most innovative and productive," he says. "So we are going to match their outside grants." Much of the money promised to competitive grants in the 2009 budget will go towards this.

Juan Pablo Paz, a physicist at the University of Buenos Aires, sees these actions as vital to giving science a greater cultural and political presence in Argentina. "Science has not been viewed as something that creates economic opportunity," he says. "But now it is being seen as part of our economic policy. I am very encouraged."

Outside the country though, there is still hesitancy. Oscar Bruno, a mathematician who attended Kirchner's New York meeting, left his homeland in 1986 to complete a doctorate in mathematics at New York University. Now he is a tenured professor at the California Institute of Technology in Pasadena. Barañao is creating an institute for applied mathematics, a public–private partnership based in Buenos Aires. But although Bruno feels the cultural draw to return, he is not yet sure if the incentives and investment in Argentinian science are enough to persuade him to leave the US lifestyle he has adopted. "I am watching and thinking," he says. "I would like to help my country. But returning is a very difficult call." ∎

Rex Dalton is a reporter for *Nature* based in San Diego.

1. *Nature* **450,** 598 (2007).
2. de la Mata, M. *et al. Mol. Cell* **12,** 525–532 (2003).
3. Swartz, T. E. *et al. Science* **317,** 1090–1093 (2007).
4. Marris, E. *Nature Med.* **13,** 1130 (2007).
5. Basu, P. *Nature Med.* **13,** 1128 (2007).
6. Toni, N. *et al. Nature Neurosci.* **11,** 901–907 (2008).

R. DALTON

C. MAZZEO/AGENCIA CYTA-LELOIR INST.

CORRESPONDENCE

What other treasures could be hidden in conference papers?

SIR — In the obituary of Anatol Zhabotinsky (*Nature* **455,** 1053; 2008), Irving Epstein mentions Boris Belousov, with whom Zhabotinsky shared the Lenin Prize in 1980 for their contributions to the Belousov–Zhabotinsky oscillatory chemical reaction system.

Epstein says "Belousov tried to publish his results in peer-reviewed journals, but eventually gave up after referees and editors insisted that such behaviour contradicted the Second Law of Thermodynamics. He instead published a one-page description of his observations in an obscure conference proceedings on radiation medicine." That paper[1], 'A periodic reaction and its mechanism', gained little attention at the time.

Papers published in symposium proceedings do not usually merit citation, because they are not peer-reviewed. They receive little recognition. Very few are even indexed in the main journal databases — one notable exception being PubMed's listing of the annual Cold Spring Harbor Symposium on Quantitative Biology.

However, other 'hidden' conference papers have also subsequently provoked acclaim. The pioneering work of physicist Abdus Salam and chemist Koichi Tanaka aroused little interest when it was first published in this way[2,3]. Fortunately, these findings were later recognized for their originality and importance: Salam went on to win the 1979 Nobel Prize in Physics, and Tanaka was awarded the 2002 Nobel Prize in Chemistry.

Min-Liang Wong Department of Veterinary Medicine, National Chung-Hsing University, Taichung 402, Taiwan
e-mail: mlwong@dragon.nchu.edu.tw

1. Belousov, B. P. *Compil. Abstr. Radiat. Med.* **147,** 145 (1959).
2. Salam, A. in *Elementary Particle Theory, Proceedings of the Nobel Symposium held in 1968 at Lerum, Sweden* (ed. Svartholm, N.) 367–377 (Almqvist & Wiksell, 1968).
3. Tanaka, K. *et al.* in *Proceedings of the Second Japan–China Joint Symposium on Mass Spectrometry* (eds Matsuda, H. and Liang X. T.) 185–188 (Bando, 1987).

Public opinion and the ethics of primate brain research

SIR — Several issues in your News story 'German authority halts primate work' (*Nature* **455,** 1159; 2008) call for clarification. Freedom of research is written into German basic law, but so is animal protection. This reflects concerns in society at large about the ethics of subjecting animals to pain and distress in research, as well as in farming and for entertainment. Even a prominent scientist such as Andreas Kreiter must justify his use of animals.

Although Kreiter refers to the ethical judgement by Bremen's senate of health as "purely arbitrary", it is backed by a political majority in the Bremen Senate as well as by the majority of Bremen's citizens, as confirmed in petitions and opinion polls.

You say that "the ruling ignores a positive judgement rendered last year by an expert commission comprising scientists and representatives of animal-welfare organizations". But the commission restricted itself to assessing the scientific merits of Kreiter's research, not the ethical issues — thereby failing in part of its mandate, which expressly included ethical issues.

Also, there was only a single animal-welfare specialist among the five members of this commission; the remainder were scientists who conduct brain research in primates or breed them for research. The previous year, one of them had himself been denied permission, on ethical grounds, to conduct invasive brain research. Moreover, you imply that the animal-welfare specialist also approved the monkey experiments. Nothing could be further from the truth.

This is not the first time that the expert group's judgement has been misrepresented to the media by scientists and university officials. Again and again the German Animal Welfare Federation has been forced to try and correct the mistaken impression that experts in science, ethics and animal welfare unanimously endorsed Kreiter's project.

Kreiter's failure to explain satisfactorily to the public exactly what he is doing is seriously undermining his credibility, and that of scientists in general. Insisting that the ethical concerns are unreasonable and that the constitutional mandate of animal protection is an undemocratic assault on academic liberty deepens the antagonism between town and gown.

You quote Stefan Treue as saying he "just can't see why what's perfectly fine in one place should be unethical in another". In fact, monkey-brain research much like Kreiter's in the level of suffering it causes has been prohibited in Munich, Berlin and Zurich.

Kreiter's centre for primate research in its present form is now in jeopardy. He should face the fact that ethical standards have evolved since he started this work and that he has lost touch with the majority of his fellow citizens.

Ulrike Gross German Animal Welfare Federation, Animal Welfare Academy, Spechtstraße 1, 85579 Neubiberg, Germany
e-mail: ulrike.gross@tierschutzakademie.de

Marker metabolites can be therapeutic targets as well

SIR — Your News & Views Q&A article 'Systems biology: metabonomics' (*Nature* **455,** 1054–1056; 2008) highlights the importance of metabonomics in the identification of metabolites associated with disease — for example, as biological markers for disease state and susceptibility, and for monitoring response to treatment. However, metabonomics can also be useful for determining the therapeutic potential of metabolites whose levels are altered in a particular disease state.

If changing concentrations of a specific metabolite can be linked to the genesis or progression of a disease, then there may be a therapeutic advantage in restoring these to normal values. This strategy has been successful, or at least promising, in many cases. For example, several anticancer treatments exploit the antiproliferative action of ceramide, the concentration of which decreases in certain cancer types (C. P. Reynolds *et al. Cancer Lett.* **206,** 169–180; 2004). Also, increasing the concentrations of *S*-nitrosothiol metabolites in the airway-lining fluid, which are lowered in patients with asthma, seems to have a protective effect in animal models (L. G. Que *et al. Science* **308,** 1618–1621; 2005).

Metabolites have a variety of cellular functions, including acting as direct regulators of gene expression, so it is not surprising that they can also function as effectors of molecular events that contribute to disease. Those positively associated with disease causation may be rarer than those that simply result from a disease.

The human metabolome comprises thousands of endogenous molecules, many of whose functions are unknown. We believe that the concept of disease-associated metabolites as potential therapeutic agents is underexploited, in comparison with their widespread use as biological markers.

Adrian K. Arakaki, Jeffrey Skolnick Center for the Study of Systems Biology, Georgia Institute of Technology, 250 14th Street NW, Atlanta, Georgia 30318, USA
e-mail: adrian.arakaki@gatech.edu
John F. McDonald School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, Georgia 30332, USA

Contributions may be submitted to correspondence@nature.com.

BOOKS & ARTS

Science, dogmas and the state

Misrepresentation of stem-cell science in Italy by political and religious groups is damaging that nation's laws and the funding and perceived value of biomedical research, argues **Elena Cattaneo**.

Staminalia: Le Cellule Etiche e i Nemici Della Ricerca
by Armando Massarenti
Guanda: 2008. 205 pp. €14.50 (in Italian)

On 1 April 2004, Elizabeth Blackburn, co-discoverer of the telomerase enzyme, published an article in the *New England Journal of Medicine* entitled 'Bioethics and the political distortion of biomedicine'. In it she reported her experience of working on, and being suddenly fired from, the US President's Council on Bioethics as it drew up its report on monitoring stem-cell research. Blackburn suggests that she was removed because her evidence-based views were not politically acceptable.

Sadly, similar cases have become common in Italy in the past two decades, particularly in the arena of stem-cell research, as Armando Massarenti explains in *Staminalia: Le Cellule Etiche e i Nemici Della Ricerca* (Staminalia: Ethical Cells and the Enemies Of Research). Massarenti, a science philosopher who writes for the Italian newspaper *Il Sole 24 Ore*, describes the political and bioethical disputes under way in Italy, highlighting the contradictions in how research carried out on stem cells derived from early human embryos (blastocysts) compares with studies on stem cells from adult tissues. He reveals the existence of a strategy to generate confusion between the scientific, ethical, religious and political aspects of stem-cell work.

Stem-cell research has evolved rapidly, from James Thomson's first isolation of human embryonic stem cells in 1998 to Shinya Yamanaka's recent discovery of cellular reprogramming. As the field has developed, researchers have held vigorous discussions about the validity and repeatability of results. In parallel, a debate has broken out on the morality of experimenting with early human embryos or their stem-cell derivatives, as some religions or moral philosophies consider human embryos consisting of more than 100 cells to be ontologically and symbolically equivalent to a human.

Most countries have avoided misinterpretation of the results and uncertainties of stem-cell research for political purposes. But in Italy manipulation has been widespread, the book reports. Indeed, some members of the Catholic church hierarchies still claim that research on embryonic stem cells is unnecessary. As

Protesters' efforts to overturn Italy's law on embryonic stem cells were thwarted by misreported science.

Staminalia describes, Yamanaka and colleagues noted the continued need for embryonic stem-cell research in a 2007 letter in *Cell Stem Cell* about the discovery of induced pluripotent stem cells from adult tissues; yet the religious media reported the opposite. Other parts of the Catholic political milieu have incorrectly stated that scientists working on embryonic stem cells in Italy are acting against the law. And I have myself been depicted in the media as a bad scientist and teacher and worse, following my organization of an open workshop on embryonic stem cells.

Such distortions coloured the law on *in vitro* fertilization passed by the Italian parliament in 2004. It made it illegal to derive embryonic stem cells from supernumerary frozen blastocysts, and the political process itself stirred up wider opposition to stem-cell research. With few exceptions, the majority of Italian biomedical scientists protested against the law in a subsequent referendum campaign. They objected to the misuse of science to sustain a controversial ethical debate with consequences for research freedom, including the risk of reducing opportunities to learn about early-derived embryonic stem cells and their developmental potential. The protest did not stop religious hierarchies and political parties from continuing their campaign. Despite a 2006 Eurobarometer survey that showed more than 60% of Italians approved of embryonic stem-cell research,

the referendum failed to overturn the law. Massarenti reminds us that the media also played a role in conveying the incorrect view of a split scientific community, as if scientists debated the ideological issues in the same way as did the political and religious personalities.

Similar attitudes have prevailed in the United States under President George W. Bush. There, a presidential decree prohibits the public financing of embryonic stem-cell research — but this is camouflage. By allowing private funding of the same research to continue, the government can pretend to ban something yet still take advantage of it. In Italy and the United States, politicians are allowing religious ideas to influence the rules of a state and opposing science without clarifying the consequences to the citizens who have elected those politicians.

The book then comes to its most crucial point. Those opposed to embryonic stem-cell research in Italy and elsewhere are not simply presenting their ethical or religious arguments and asking those who share them to adopt a consistent behaviour. Rather, they are denigrating scientific results by emphasizing disagreements and spreading false information about the alleged scientific or therapeutic superiority of the research that they wish to support. This approach is applied to stem-cell research today, but tomorrow could be directed at any other field of science judged to be troublesome.

Misinformation has consequences for the

political guidelines that sustain research. In countries where funding allocations are based on peer review, these effects should be containable. Competition for the best ideas will not depend on a scientist's political or religious points of view or public perception. Where conflicts of interest pollute the management and public funding of science, as in Italy, misinformation may inspire and strengthen political interference with devastating effect, beyond damaging the research that could otherwise enhance the cultural and economic contribution of a country rich in creativity.

Where a peer-review culture is lacking, a country will not benefit from open competition for the best ideas. Negotiation between the public administration and research institutions will lead to an allocation of funds without a transparent evaluation system. Committees will finance their own members, and financial acts could assign millions of euros of public money every year to favoured institutions, according to the biases of ministers.

Such a background of patronage is fostering the political distortion of biomedical research and undermining the autonomy of science in Italy, in my view. Factors leading to the manipulation and censoring of Italian science, particularly stem-cell research and plant biotechnologies, should be analysed thoroughly. This would educate the public in the ethics of science and prevent similar situations from arising in other countries.

Italy does have excellent scientists who understand the basic concept of peer review. What is missing is ethical public-asset management. As Massarenti explains, many researchers find it difficult to accept that it is impossible to support and invest in science effectively if conflicts of interest are not eliminated. Other countries have recognized this. In 1986 Spain created the ANEP, a national

evaluation and foresight agency that is charged with selecting the best projects on behalf of ministers, regions and even private foundations. I hope that a similar mechanism will be adopted in Italy for public funding of biomedical research. The Italian government is moving in the right direction, through propositions from some of its past and current members. But a reliable peer-review system requires two things: the removal of conflicts of interest and of the suspicion that decisions may be influenced by personal interests.

The ethical dimension of science is at the core of *Staminalia*. Science's objectivity has been appreciated since Galileo Galilei, even though for a long time it was judged as a heresy. Yet, many principles of democratic coexistence have been built on this ethical dimension. The possibility of criticizing and reviewing results is an essential part of science and of the moral and civil growth of a nation. Those attacking these values and representing science and scientists as a threat to humanity are expressing intolerance and contempt for democracy itself. ■

Elena Cattaneo is a professor in the Department of Pharmacological Sciences and director of the Centre for Stem Cell Research, University of Milan, Milan, Italy.
e-mail: elena.cattaneo@unimi.it

Golden eggs symbolizing fertility were used in a campaign against Italy's *in vitro* fertilization law.

A blogging professor in print

Academeology: Random Musings, Strong Opinions and Somewhat Bizarre Anecdotes from an Academic Life
by Female Science Professor
Lulu: 2008. 283 pp. £9.26

As a science undergraduate, I imagined that being a professor involved little more than teaching a class each semester and 'doing research'. The inner workings of academia were a mystery to me. Since then, an explosion of blogs written by students, postdocs and faculty members has opened a new window into academic life. Blogs allow scientists at all levels to peek into the professional lives of colleagues at other institutions and in different fields.

Pseudonymous blogger Female Science Professor (FSP) first introduced herself in May 2006: "I do not look my age, I do not look like a professor, I do not look like a scientist. My colleagues are, with a few exceptions, very kind and polite to me, and some (many? most?) even like me … but they do not take me seriously."

That post set the tone for what has become a collection of more than 500 short essays describing her experiences as a scientist, a professor and one of the few women working in her field of science. She has revealed little about her life beyond the fact that she is a 40-something professor in a physical-science department at a large US university, is married to a fellow science professor and is mother to a "tween"-aged daughter. FSP's blog has steadily gained popularity, with her clear writing style, candid revelations and often humorous musings. The blog's comment section allows her readers to share their own academic experiences.

FSP has published a selection of her essays as a book, *Academeology*. Arranged by theme, the posts are interspersed with pithy points

of academic etiquette. Many of the topics she covers, such as getting a job in academia, tips on teaching, giving seminars and writing grants, might be found in any academic career guide. But rather than give generic advice, FSP presents her own experiences in an informal and entertaining style. Writing under a pseudonym allows for candid descriptions of students and colleagues. Among the topics she covers are graduate student admissions from the perspective of a faculty member, dealing with coauthors — and advice on how not to be a 'sexist jerk'.

FSP's stories of being a woman in a male-dominated field are engrossing. She describes the casual sexism, such as being ignored or treated as a secretary by visiting scientists, or having male colleagues comment that she received an award "because she is a woman". These tales might be disheartening to some. But FSP also relates her successes as a scientist and in navigating difficulties as one half of a scientist couple who began her academic career with a young child. Never claiming that her experiences are universal or that her path has been easy, FSP shows that it is possible to have both a career as a scientist and a life outside of science.

Yet the book's strength in discussing many aspects of academic life is also a weakness. Each essay has a different audience: some are aimed at undergraduates applying to graduate school, others at graduate students or fellow faculty members. The style and tone vary between entries. And the book misses the added value of the comments associated with FSP's original blog entries. That said, *Academeology* is an enjoyable read for anyone interested in how academia works. I would have loved to have had a copy during my own student days. ■

Peggy Kolm is a biotechnology consultant based in California, and blogs at sciencewomen.com.
e-mail: peggy.kolm@gmail.com

Imprisoned by intelligence

The divide between science and society is extrapolated to the extreme in Neal Stephenson's novel *Anathem*. The author, who is well regarded for his vision of science in contemporary and historical settings, creates in his latest work the fictional planet Arbre, which parallels Earth in the very far future.

Following a series of past technological catastrophes for which they were blamed, the planet's scientists, scholars and philosophers — known as the avout — have been forced to live apart from the 'Sæcular world' of society at large. The avout inhabit cloistered compounds, the distinctive clock towers of which are familiar features in most large towns and cities on Arbre. Most high technology is withheld from the avout in case they repeat past disasters. Instead, they spend their robed, austere days engaged in theoretical research and low-tech experimentation using permitted objects, such as telescopes. The avout are not allowed to breed, lest their offspring harbour dangerous levels of intelligence. Rather, they repopulate their numbers with unwanted orphans and those with abilities.

As the avout go about their rituals, the Sæcular world might as well be another planet: its inhabitants chatter on 'jeejahs' (mobile phones) and surf the 'Reticulum' (Internet) when not cheering on sports teams and growing obese on sugary drinks. The two worlds mingle only in strictly controlled circumstances, such as at an annual goodwill festival and in avout-run universities where smarter Sæculars can get a sanitized education. Sæcular artisans may refine existing technology but lack the advanced scientific education to be innovative, so the avout are not dispensable. But it takes an extreme global problem for the Sæcular government to summon the avout to contribute their brain power.

So begins the novel. *Anathem* centres on a young avout, Fraa Erasmas, who comes of age during a pivotal time in Arbre's history. A starship is orbiting the planet, and its crew, although familiar, do not seem to be friendly. Erasmas and his colleagues are called on to help understand the extraterrestrial threat and fend it off.

The pleasure of reading *Anathem* derives in part from viewing our own world through the distorted fairground mirror that Stephenson establishes. The language is English with a twist: the word devout morphs to *avout*, friar to *fraa*, sister to *suur*. Euphemistic management speak is *bullshyte* and a caste of scathing computer technicians is known as the *Ita*. The distortions are reflected in scientific concepts as well: Occam's razor becomes Gardan's Steelyard. And the great scientific figures of Earth get their Arbre avatars: Plato appears as Protas, Socrates as Thelenes and Archimedes as Carta. As with any decent distortion, the author leaves unexplained many areas of Arbrean society and history that the reader can enjoy filling in.

The story of Erasmas and his friends is an epic adventure that is well paced and exciting. Yet the action is underpinned by serious, carefully researched scientific and philosophical concepts — notably quantum mechanics, parallel universes and the nature of consciousness. Some of the scholarly dialogues take work to follow, but the effort is rewarded. Stephenson credits a number of philosophical and mathematical thinkers in helping him flesh out these fictionalized ideas, including Gottfried Leibniz, Kurt Gödel, Edmund Husserl and Edward Zalta. As the people of Arbre struggle to understand the alien invaders, the author seems most

In Retrospect: Fernel's *Physiologia*

The seeds of systems biology were sown 450 years ago by a classic medical text, explains **G. Rickey Welch**.

The Physiome Project, spearheaded by the International Union of Physiological Sciences, aims to generate a framework for understanding the workings of the entire human organism, from genes and proteins to the whole body. A seminal Latin medical text presaged this venture around 450 years ago.

Its author, Jean Fernel (pictured, right), was one of the foremost figures of French Renaissance science and medicine, and the first to use the term 'physiology' in its modern scientific sense. At the time, 'physiologia' denoted the study of nature or natural philosophy. First published in 1542, *Physiologia* underwent more than 30 reissues, and caused debate in Europe for the next 100 years.

Physiologia, together with *Pathologia* and *Therapeutice*, was one of three volumes that comprised Fernel's crowning work, *Universa Medicina*. *Physiologia* set forth a fundamental theory of the operation of the human body: the nature of the wholly healthy human being, all the powers and functions. This flowed into a description of diseases (*Pathologia*) and their treatments (*Therapeutice*). Fernel's opus resonates today in the goal of the Physiome Project, which is 'to understand and describe the human organism, its physiology and pathophysiology, and to use this understanding to improve human health'.

In Fernel's time, medical practice centred on the study of body parts in lavishly illustrated anatomy texts, by authors such as Andreas Vesalius. Fernel attacked this 'shallow and easy art', arguing that it lacked holistic principles of causation and function. His answer, *Physiologia*, is a brilliantly logical treatise. Fernel begins with a philosophical reflection on the appropriate means of analytical attack, and he settles on a top-down method — one that we recognize in today's physiological practice. He then describes all the known anatomical parts of the human organism and develops his notion of a 'comprehensive physiology.' Seizing on the Aristotelian principles of deductive reasoning, causal analysis and physics of matter, Fernel demonstrates how the elemental attributes of the body parts are woven into a whole that manifests all the temperaments, humours, powers and faculties of the living being. Although he characterizes the causation of bodily processes within the framework of the 'animal spirits' proposed by the ancient Greek physician Galen, Fernel envisaged such 'spirits' in a real material form, breaking from the irrationality of occultism and magic that had dominated medieval natural philosophy and medicine.

Physiologia constitutes the first complete
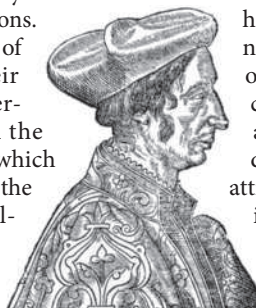
WELLCOME LIBRARY, LONDON

indebted to Roger Penrose's controversial ideas about how the quantum world might impinge on the neurobiology of consciousness.

As a thought experiment in examining the relationship between science and society, and as a cautionary tale about the consequences of its breakdown, *Anathem* provides much to ponder. In one scene, a scholar quizzes her students about the various stereotypes of the avout that society has harboured. The images are familiar: scientists as loveable, dishevelled Einsteinian boffins; as mystics hoarding the Universe's secrets; as criminally insane desperadoes in white smocks with schemes to take over the world; as highly strung, meddling know-it-alls who simply don't understand the realities of modern life. Although details of the catastrophes that caused the avout to be sequestered are not made clear, it is hinted that tinkering with genetic information, with the structures of atoms and with space-time itself were all factors. *Anathem* is a shrewd exploration of what might happen if the fear of scientists meddling with things they were never meant to know becomes entirely justified. ∎

Jennifer Rohn is a cell biologist at University College London, London WC1E 6BT, UK, editor of LabLit.com and author of *Experimental Heart*.
e-mail: jenny@lablit.com

view of what we now call physiology. It is also a forerunner of the systems approach to understanding form and function in living organisms. The field of systems biology today presents us with a challenge similar to that faced by Fernel — to interpret staggering amounts of data depicted in numerous pictographic forms, such as DNA microarray snapshots, proteome network displays and protein–protein interaction maps. Although causation has come to be assigned to a molecule, namely DNA, the way in which we should map from genes to the organism is far from apparent. The Physiome Project embraces the systems-biology movement by showing us how the whole is more than the sum of the parts.

Denis Noble, one of the founders of the project, proffered that "a major part of the future of physiology surely lies in returning to our roots. Higher-level systems biology is, I suggest, classical physiology by another name." We might say that the Physiome Project began at the dawn of the scientific revolution with *Physiologia*. Fernel's Renaissance thinking pointed the way to a 'universal medicine' that we are now poised to attain. ∎

G. Rickey Welch is professor in the departments of biological sciences and history at the University of Maryland, Baltimore, Maryland 21250, USA.
e-mail: welch@umbc.edu

Mark Rothko's *Red on Maroon*.

Rothko's methods revealed

Rothko: the Late Series
Tate Modern, London
Until 1 February 2009

Mark Rothko's Seagram murals, commissioned in 1958 to decorate the Four Seasons restaurant in the Seagram Building in New York, are famous for their layers of translucent and opaque paints that produce a luminous and ephemeral quality.

Rothko was notoriously secretive about his methods, refusing to let even his studio assistants watch him paint. Yet the clues he left within the murals have informed conservators of the variety of materials he used and his layering techniques that were innovative for the time.

Conservators at the Tate Modern have studied the murals using cutting-edge techniques in collaboration with researchers at MOLAB, an Italian organization that provides technical support to European conservation projects. They investigated the chemistry of microgram samples of paint using mass spectrometry, and probed the structure of layers with high-resolution electron microscopy. Ultraviolet images reveal Rothko's brush work, each layer made from a unique medium that fluoresces differently.

Their research shows that Rothko used materials far beyond the conventional range sold for artists, modifying the properties of oil paints to achieve the flow, drying time and colours he needed. He used synthetic substances such as oil-modified alkyd and acrylic resins alongside traditional materials, including egg, glue and dammar resin, which are fast-drying and allowed him to apply subsequent layers within hours. Resins increased the viscosity of the mixtures so the paints could be diluted without losing their coherence. Rothko also applied phenol formaldehyde to prevent layers from blending into one another. Each mural differs with regard to its paint mixture or the layering sequence, suggesting that Rothko constantly experimented.

The paintings are fragile, prone to damage by handling, and various components in the layers may react and age differently. Rothko chose not to apply varnish, which would have destroyed his subtle effects of gloss and matt variation, and he did not want to exhibit his creations framed behind glass. Knowledge of the materials present in the murals will help conservators develop bespoke techniques to protect them for future generations. ∎

Jane Qiu is a writer based in Beijing and London.
e-mail: jane@janeqiu.com

PARTICLE PHYSICS

Mass by numbers

Frank Wilczek

A highly precise calculation of the masses of strongly interacting particles, based on fundamental theory, is testament to the age-old verity that physical reality embodies simple mathematical laws.

In a milestone paper, Dürr *et al.*[1] report a first-principles calculation of the masses of strongly interacting particles (hadrons, such as the proton), starting from the basic equations for their constituent particles (quarks and gluons), and including carefully documented estimates of all sources of error. Their results, published in *Science*, highlight a remarkable correspondence between the ideal mathematics of symmetry and the observed reality of the physical world.

Quantum chromodynamics (QCD), the theory of the so-called strong force or strong interaction, postulates elegant equations for quarks and gluons. Those equations embody enormous symmetry, which largely dictates their form. A dramatic reflection of this conceptual rigour is that the equations contain very few freely disposable parameters — just a mass for each 'flavour' of quark (u, d, s, c, b, t) and an overall coupling constant. This makes QCD, in principle, an extremely powerful predictive framework. In fact, it's even tighter than this accounting suggests: for many purposes one can ignore the heavy quarks (c, b, t) and absorb the coupling constant into an overall scaling factor.

QCD predicts, however, that quarks and gluons are not observable particles. Rather, they occur only as building blocks inside more complex objects, collectively dubbed hadrons. The most familiar hadrons are protons and neutrons, from which ordinary atomic nuclei are assembled. Over decades of investigation, dozens of additional hadrons have been discovered. Most of these relatives of protons and neutrons are highly unstable, but their properties — notably their mass, charge and spin — can be measured[2].

If QCD is valid, its equations should account for all the properties of hadrons. But it has proved extremely challenging to solve those equations with enough precision to enable a sharp, quantitative comparison between
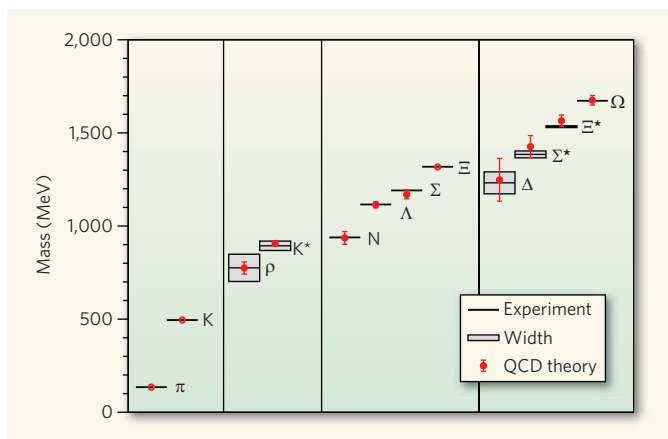
Figure 1 | Theory meets experiment. The masses of light hadrons (expressed in energy units; 1 MeV corresponds to 10^6 electronvolts) computed by Dürr *et al.*[1] using quantum chromodynamics (QCD) calculations (filled circles) are in remarkable agreement with the experimental values (horizontal lines). Each symbol (π, K and so on) refers to a different type of hadron, according to quark complement, spin and so on. The widths of the bands indicate the experimental decay widths, which are related to the finite lifetime of the particles. The vertical error bars denote the theoretical error estimates. Three of the hadrons (π, K and Ξ) have no error bars because they are used to fix the theory's parameters. The vertical lines divide the particles into four groups according to spin: 0, 1, $\frac{1}{2}$, $\frac{3}{2}$, reading from left to right. (Modified from ref.1.)

theoretical predictions for, and experimental measurements of, hadron properties. The full power of modern, massively parallel computing has been brought to bear on this problem. Several impressive partial results have been announced in recent years[3]. Now Dürr *et al.*[1] have assembled all the pieces systematically, and added some new refinements, to achieve a fully convincing, successful comparison at a level of precision of 1–2% (Fig. 1).

A key aspect of their calculations is the estimation of errors. We know the equations of QCD precisely, but practical calculations require several approximations. To appreciate these approximations, we must first briefly review some specific, unusual features of the equations and their solution.

The primary objects in the theory of QCD are quantum fields — the quark and gluon fields. Quantum fields are entities that fill all space and exhibit spontaneous activity. That spontaneous activity is often referred to as

quantum fluctuations, or 'virtual particles'. In the mathematical formulation, there is a master wavefunction for the quantum fields. The wavefunction is a superposition of different possible patterns of excitation in the fields, each occurring with some definite amplitude. The central problem involved in solving the equations of QCD, to predict the census of hadrons and their properties, is to compute this wavefunction: that is, to determine the numerical value of the amplitudes. Having constructed the wavefunction of 'empty space', we can inject different combinations of quarks and gluons and study the equilibria they settle into. Those equilibria correspond to observable particles — the hadrons.

The possible patterns of excitation in continuous quark and gluon fields map out a space of infinite dimensionality — roughly speaking, we should have to specify 84 $((3 \times 3 \times 4) + (8 \times 6))$ numbers at each point in space. For the quark fields there are three flavours, three colours, and four components accounting for spin and antiparticles; for the gluon fields there are eight directions in the space of its symmetry group, and for each direction there are six fields: three electric and three magnetic. In principle, we should calculate the amplitude of each such pattern. But no computer can handle an infinite number of variables, so two types of approximation seem unavoidable: the spatial continuum must be replaced by a discrete lattice of points; and the calculations must focus on a finite volume.

The process of discretization, which might seem to be a drastic mutilation of the theory, is actually well controlled theoretically, owing to QCD's central property of asymptotic freedom[4]. In this context, asymptotic freedom implies that the short-wavelength fluctuations of the fields, which are the ones we lose track of when we discretize space, are of a very simple form. In technical jargon they approach Gaussian

random fields, or what physicists call 'free fields'. Thus the effects of the missing fluctuations can be computed analytically and added back in.

The approximation of finite volume is mitigated by the fact that in QCD the fundamental interactions occur among field variables at neighbouring points (we say they are local interactions). The patterns of equilibrium that define hadrons are, however, generally spatially extended, and so it is important to take a large enough volume so they fit comfortably. It is possible to control finite-volume errors by varying the simulated volume and making theoretically informed extrapolations.

Two additional approximations have also proved unavoidable, and troublesome, in practice. One is that as the u and d quark masses are taken down to their (very small) physical values, the equations get harder to solve. (For experts: this is because there are long correlation lengths, and the equations become numerically 'stiff'.) Like the compromise of assuming a finite volume, this is handled by sophisticated, theoretically informed extrapolation from simulations using larger mass values. Finally, even after acceptable levels of discretization and restriction to finite volume, the space that should be surveyed by the wavefunction is far too large for even the most powerful modern computer banks to handle. So in place of a complete survey, we must content ourselves with a statistical sample of the wavefunction. This introduces errors that can be estimated by the standard techniques of statistics.

For optimal use of resources, one should bring all the important sources of error to the same level. This involves a delicate balancing act. For example, using larger volumes or smaller quark masses requires lengthier calculations, which degrade the sampling rate of the wavefunction. The technical feat of Dürr et al.[1] is to achieve such a balance, keeping all the errors demonstrably small.

Of course, overwhelming evidence for the validity of QCD has been accumulating for decades, from very different sorts of calculations and experiments. Although quarks and gluons do not exist as isolated particles, they can be reconstructed from the patterns of energy–momentum flow they imprint on hadrons. In high-energy collisions, the emerging hadrons are found to be organized into jets of particles moving in approximately the same direction as each other. According to QCD, if we replace the jets by fictitious single particles with the same total energy and momentum as the jets, those fictitious particles will obey the equations of elementary quarks and gluons. This is another aspect of asymptotic freedom. Through the study of jets, the basic equations of QCD have been verified in exquisite detail.

So what value is added by using already-validated equations to compute already-measured hadron masses? One answer is practical. The same techniques that are used to compute known hadron masses can also be used to

compute other interesting quantities that are very difficult to measure experimentally. For example, some key reactions involving small nuclei and unstable particles (hyperons) are very important in stellar nucleosynthesis and supernova dynamics, but are impracticable to measure. Having numerical techniques that reliably reproduce what is known, we can address the unknown confidently.

But perhaps a more profound answer is philosophical. A great vision of science — stretching from Pythagoras' credo "All things are number", to Kepler's ordering of the planets based on Platonic solids, to Wheeler's slogan "Its from bits" — has been that physical reality embodies ideally simple mathematical laws. As physics developed before the quantum revolutions of the twentieth century, the basic equations emphasized dynamics (how given systems evolve in time) as opposed to ontology (the science of what exists). Kepler's system was stillborn, but in the world of QCD and hadrons, the great vision lives and thrives.

Finally, let me add a note of critical perspective. The accurate, controlled calculation of hadron masses is a notable milestone. But the fact that it has taken decades to reach this milestone, and that even today it marks the frontier of ingenuity and computer power, emphasizes the limitations of existing methodology and challenges us to develop more powerful

techniques. QCD is far from being the only area in which the challenge of solving known quantum equations accurately is crucial. Large parts of chemistry and materials science pose similar mathematical challenges. There have been some remarkable recent developments in the simulation of quantum many-body systems, using essentially new techniques[5]. Can the new methods be brought to bear on QCD? In any case, it seems likely that future progress on these various fronts will benefit from cross-fertilization. The consequences could be enormous. To quote Richard Feynman[6]: "Today we cannot see whether Schrödinger's equation contains frogs, musical composers, or morality — or whether it does not. We cannot say whether something beyond it like God is needed, or not. And so we can all hold strong opinions either way." ■

Frank Wilczek is in the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA.
e-mail: wilczek@mit.edu

1. Dürr, S. et al. Science 322, 1224–1227 (2008).
2. http://pdg.lbl.gov
3. Aubin, C. et al. Phys. Rev. D 70, 094505 (2004).
4. Wilczek, F. in Les Prix Nobel 100–124 (Almqvist & Wiesell Int., 2004).
5. Verstraete, F. & Cirac, J. I. Preprint at http://arxiv.org/abs/cond-mat/0407066 (2004).
6. Feynman, R., Leighton, R. & Sands, M. in The Feynman Lectures on Physics Vol. 2, Ch. 41, 12 (Addison-Wesley, 1964).

Turtle origins out to sea

Robert R. Reisz and Jason J. Head

Various aspects of turtle evolution are the subject of vigorous debate among vertebrate palaeontologists. A newly described fossil species, the oldest yet discovered, adds grist to the mill.

During the Late Triassic, some 220 million years ago, primitive turtles about 40 centimetres in length were preserved in sedimentary deposits in what is now southwestern China. These fossils are examples of a new species of a very early turtle, named *Odontochelys semitestacea*, which is described by Li et al. on page 497 of this issue[1] and which will change ideas about turtle origins and the evolution of their striking body plan.

Turtles are remarkable animals[2]. They have a horny beak rather than teeth, and a shell like that of no other animal, one that is composed of an upper carapace and a lower plastron, jointed together by a bony bridge. The shell is a composite structure derived from ribs, parts of the shoulder girdle and specialized dermal bones. This precludes the typical costal respiration of tetrapods, in which movable ribs allow the chest cavity to expand and contract. Turtles have overcome this obstacle by having the muscles that control breathing use the

limb pockets at the borders of the shell. This shell has become modified as turtles diversified and adapted to terrestrial, amphibious and aquatic environments (Fig. 1). The evolutionary relationships and ecology of turtles through time, and the developmental and evolutionary origins of the shell, are major controversies in studies of vertebrate evolution.

Previously, the fossil evidence for turtle origins came largely from *Proganochelys quenstedti* from Germany, which lived between 204 million and 206 million years ago, and other less-well-known early turtles. *Proganochelys* is known from several skeletons. It has a massive shell and spiked armour on the neck and tail, but also retains teeth on the roof of the mouth and has other primitive features in the skull and skeleton. Its osteology has been used to propose[3] that turtles are related to pareiasaurs, a group of extinct parareptiles that includes species with extensive dermal armour. And on the basis of evidence from *Proganochelys*

R. R. REISZ

Figure 1 | Differences in turtle ecology. There are about 300 extant species of turtle, which vary in weight from about 100 grams to about 860 kilograms, and whose habitat ranges from wet meadows to deserts, and from rivers and lakes to the open ocean. Top, a fully terrestrial turtle (*Gopherus agassizii*), representing a similar ecology to that of the primitive turtle *Proganochelys* and its relatives inferred in previous studies[4,5]. Bottom, an aquatic turtle (*Apalone spinifera*), representing a potentially similar ecology to that of the fossil species *Odontochelys semitestacea* described by Li and colleagues[1].

and its close relatives from the Late Triassic of Argentina, Germany and North America, it has been suggested that the earliest turtles lived in terrestrial environments[4,5].

The discovery and description of *Odontochelys* by Li *et al.*[1] challenges these hypotheses. *Odontochelys* is not only the oldest recognizable turtle, but its skull also shows that it is more primitive than other turtles because it retains a full complement of marginal teeth, rather than a beak, and also possesses free sacral ribs and a long tail. Its osteology contradicts the view that turtles have pareiasaurian affinities, and, along with molecular data, supports evolutionary hypotheses that they are closely related to another group, the diapsid reptiles[6,7]. Li *et al.* argue that *Odontochelys* represents an early stage in the evolution of the turtle shell because the plastron is present and fully developed, but the carapace is apparently absent, with only dorsal ribs and neural (midline) dermal ossifications present. The authors infer, therefore, that the plastron evolved before the carapace, reflecting the timing of shell ossification during embryonic development in living turtles.

Although this evolutionary scenario is plausible, we are particularly excited by an alternative interpretation and its evolutionary consequences. We interpret the condition seen in *Odontochelys* differently — that a carapace was present, but some of its dermal components were not ossified. The carapace forms during embryonic development when the dorsal ribs grow laterally into a structure called the carapacial ridge, a thickened ectodermal layer unique to turtles[8]. The presence of long, expanded ribs, a component of the carapace of all turtles, indicates that the controlling developmental tissue responsible for the formation of the turtle carapace was already present in *Odontochelys*. The expanded lateral bridge that connects the plastron to the carapace in other turtles is also present, implying that the

plastron was connected to the laterally expanded carapace. Thus, an alternative interpretation is that the apparent reduction of the carapace in *Odontochelys* resulted from lack of ossification of some of its dermal components, but that a carapace was indeed present.

This interpretation of *Odontochelys* leads us to the possibility that its shell morphology is not primitive, but is instead a specialized adaptation. Reduction of dermal components of the shell in aquatic turtles is common: soft-shelled turtles have a greatly reduced bony shell and have lost the dermal peripheral elements of the carapace. Sea turtles and snapping turtles have greatly reduced ossification of the dermal components of the carapace, a condition similar to that seen in *Odontochelys*.

From the geological context of their fossils, Li *et al.*[1] conclude that *Odontochelys* lived in a shallow marine environment. That, combined

with similarities between its carapace and the reduced shells of modern aquatic turtles, leads us to propose that the absence of most of the dermal carapace in *Odontochelys* is a secondary loss associated with aquatic habits rather than a primitive condition, as inferred by Li and colleagues. Given the similarities between its shell morphology and early growth stages in living turtles, a simple truncation of carapace ossification, in which the adults retained juvenile features (paedomorphosis), could have been a developmental mechanism in the evolution of the reduced carapace.

Regardless of the primitive or derived nature of its shell, *Odontochelys* is in evolutionary terms the most 'basal' turtle yet found. Its discovery opens a new chapter in the study of the origins and early history of these fascinating reptiles. Both interpretations alter our views of turtle evolution: *Odontochelys* either represents the primitive ecology for turtles, consistent with the hypothesis that the turtles' shell evolved in aquatic environments[7], or it represents the earliest turtle radiation from terrestrial environments into marine habitats. Either way, these ancient turtles demonstrate yet again the value of new fossil discoveries in changing our understanding of vertebrate history. ∎

Robert R. Reisz and Jason J. Head are in the Department of Biology, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada.
e-mail: robert.reisz@utoronto.ca

1. Li, C., Wu, X.-C., Rieppel, O., Wang, L.-T. & Zhao, L.-J. *Nature* **456,** 497–501 (2008).
2. Gaffney, E. S. & Meylan, P. A. in *The Phylogeny and Classification of the Tetrapods Vol. 1: Amphibians, Reptiles, Birds* (ed. Benton, M. J.) 157–219 (Clarendon, 1988).
3. Lee, M. S. Y. *Science* **261,** 1716–1720 (1993).
4. Joyce, W. G. & Gauthier, J. A. *Proc. R. Soc. Lond. B* **271,** 1–5 (2003).
5. Scheyer, T. M. & Sander, P. M. *Proc. R. Soc. Lond. B* **274,** 1885–1893 (2007).
6. Hedges, S. B. & Poling, L. L. *Science* **283,** 998–1001 (1999).
7. Rieppel, O. & Reisz, R. R. *Annu. Rev. Ecol. Syst.* **30,** 1–22 (1999).
8. Gilbert, S. F. *et al. Evol. Dev.* **3,** 47–58 (2001).

ORGANIC CHEMISTRY

Short cuts to complexity

André B. Charette

The credit crunch is forcing people to tighten their belts, but chemists have long known the benefits of being economical with atoms. The latest synthesis of an anticancer agent shows how effective parsimony can be.

Nature produces an almost infinite number of structurally complex organic compounds that have fascinating — and potentially useful — biological properties. This has inspired generations of synthetic chemists to make not only the naturally occurring compounds, but also structurally modified analogues that have tailored properties and functions. Marine organisms are a particularly rich source of natural

products, but so far only one such class of compound has entered clinical trials: bryostatins, which have anticancer activity *in vivo*[1]. Reporting in this issue (page 485), Trost and Dong[2] describe a synthetic route to a particular bryostatin — bryostatin 16 — that drastically reduces the longest sequence of consecutive steps from the previous best of 40 down to a much more manageable 26. This might open

50 YEARS AGO

In an address to the School Broadcasting Council on November 7, Sir Ian Jacob, director-general of the British Broadcasting Corporation, referred particularly to the dependence of Great Britain, and possibly its survival, upon the widest possible diffusion of scientific skills and knowledge. He suggested that the British Broadcasting Corporation might be able to make, in its own way, "a new and massive contribution to the understanding of science in the secondary modern schools, where the need perhaps is greatest and where the shortage of good teachers is likely to be most acute".
From *Nature* 29 November 1958.

100 YEARS AGO

I have just acquired for the Canterbury Museum the skeleton of a huge blue whale (*Balaenoptera sibbaldii*). The whale was cast on to the beach at Okarito, on the west coast of the South Island of New Zealand, early this year, and measured 87 feet in length. My statement that the Okarito whale is among the largest known has been freely challenged in the local Press ... I have naturally sought information as to the length of skeletons of great whales preserved in museums, but have been unable to obtain satisfactory data. I shall be pleased, therefore, if directors of museums possessing the skeletons of large whales will kindly communicate with me direct, or, as the matter is one of general interest, through the medium of NATURE.
ALSO:
Students of the occult will welcome the elaborate paper by Dr. W. L. Hildburgh in the current issue of the Journal of the Royal Anthropological Institute on Sinhalese magic. He illustrates with copious detail the equipment of the magician, devil-dancer, and astrologist, describes their methods, and provides an ample supply of curious charms, amulets, and horoscopes. He does not enter upon the question of the origin of this system of magic.
From *Nature* 26 November 1908.

the door to a practical process for preparing the compound, and provide ready access to bryostatin analogues for drug discovery.

The bryostatin family of compounds is derived from the bryozoan *Bugula neritina*[3] (Fig. 1), and encompasses more than 20 structurally related natural products. Not only do these compounds exhibit anticancer properties, but they have also been shown to improve cognition and enhance memory in animals[4], making them interesting leads for drug-discovery efforts targeting Alzheimer's disease. But the concentration of bryostatins in *B. neritina* is low, so extraction from bryozoans is not a viable means of producing these compounds — at least, not in quantities that would allow a complete evaluation of their biological profiles.

Bryostatins have therefore long been synthetic targets of choice for chemists — not only because of the need to find a practical way of making them, but also because the structural complexity of the compounds provides a perfect opportunity to test new synthetic methods. Despite intensive efforts, only three total syntheses of bryostatins have been reported[5–7]. These achievements were rightly hailed as landmarks in organic chemistry, but they required lengthy sequences of consecutive synthetic steps (at least 40), making them impractical for the preparation of more than milligram quantities of material.

So what is it that makes bryostatins such interesting synthetic targets? A rule of thumb in synthetic chemistry is that the more chemical groups are squeezed into a small molecule, the more difficult that molecule will be to prepare. In this respect, the core structure of bryostatin features three rings (designated A, B and C; Fig. 2a) that are densely populated with chemical groups. Furthermore, two of these rings, B and C, contain motifs known as *exo*-cyclic trisubstituted alkenes (alkenes are distinguished by carbon–carbon double bonds). In the previous syntheses of bryostatins[5–7], controlling the geometry of the groups attached to these alkenes was a truly formidable challenge, for which ingenious solutions had to be invented. Equally difficult to prepare, on the basis of the experiences of the previous syntheses, is another alkene group in the molecule, which forms part of the linker between rings B and C.

Multi-step syntheses of complex molecular targets will be efficient and practical only if as many of the required reactions as possible are 'atom economical'. In an optimal atom-economical process, all of the atoms in the reactants end up in the desired product. This idea is one of twelve accepted criteria used to quantify how environmentally friendly chemical reactions are, and has been fully implemented by Trost and Dong[2] in their synthesis of bryostatin 16. The authors deliberately chose this molecule as their target because it

could serve as a precursor to produce many other bryostatins. Their strategy relied on coupling three fragments of the molecule at a late stage in the synthesis.

They made the B ring using a spectacular ruthenium-catalysed coupling reaction[8] between an alkene and an alkyne (a compound that contains a carbon–carbon triple bond). Although both fragments contain an array of potentially reactive chemical groups, the ruthenium catalyst binds selectively to the alkene and the alkyne, so inducing the formation of a new carbon–carbon bond (Fig. 2b). Such chemoselectivity is essential to obtaining efficient synthetic routes because it avoids the use of protecting groups — chemical groups that are attached temporarily to reactive parts of molecules to prevent them from interfering in desired reactions. Not only does the ruthenium-catalysed reaction form the B ring with the correct three-dimensional arrangement of substituents, it also controls the geometry of the potentially troublesome *exo*-cyclic alkene. Furthermore, the reaction is perfectly atom-economical: all of the atoms found in the starting materials (bar the catalyst) are present in the coupled product.

Trost and Dong[2] prepared the C ring in a two-step process involving two different transition-metal catalysts. The first step is a palladium-catalysed coupling[9] of two alkynes (Fig. 2c), which not only sets the scene for the formation of the C ring, but also forms the characteristic large ring present in bryostatins. This reaction is remarkable, both because it is perfectly atom-economical and because it is the first demonstration that such a large ring structure (containing 22 atoms) can be made using this kind of carbon–carbon bond-forming reaction. Such 'macrocyclizations' are notoriously problematic, so Trost and Dong have effectively added a useful entry to the existing roster of possible reactions.

The authors completed the synthesis of the C ring using a gold-catalysed reaction to form a carbon–oxygen bond between an alcohol group (OH) and the alkyne produced in the previous step (Fig. 2d). Gold catalysts activate alkynes selectively[10] in the presence of a wide variety of other chemical groups, a useful property that is used here to great effect.
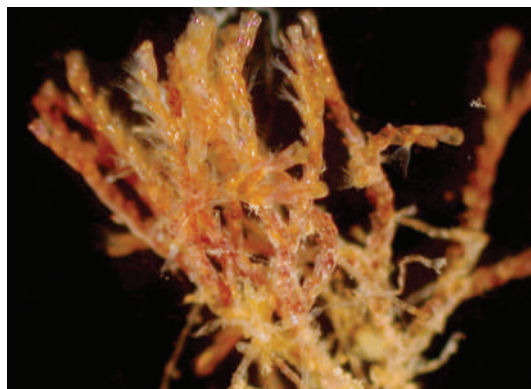

Figure 1 | The natural source of bryostatins. Approximately 1 tonne of the marine organism *Bugula neritina* is needed to isolate 1 gram of bryostatin anticancer agents.

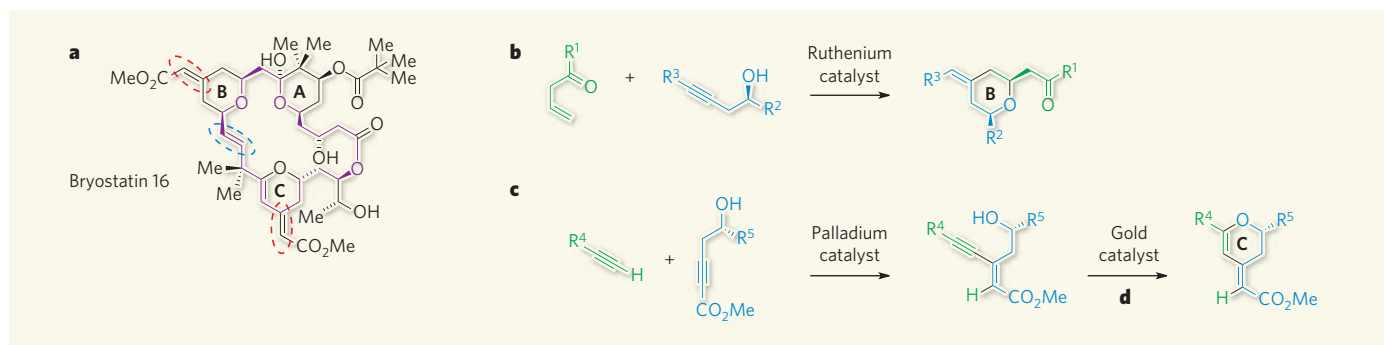SOUTHEASTERN REGIONAL TAXONOMIC CENTER/ SOUTH CAROLINA DEPT NAT. RES.

Figure 2 | Catalytic steps in the synthesis of bryostatin 16. **a**, Bryostatin 16 has been prepared in a remarkably short synthesis by Trost and Dong[2]. The molecule contains several rings, including three designated A, B and C. The large ring highlighted in purple was constructed in the palladium-catalysed reaction shown in **c**. The structure also has two *exo*-cyclic trisubstituted alkenes (circled in red) and a disubstituted *trans*-alkene (circled in blue) that are difficult to prepare. The authors used three pivotal transition-metal-catalysed steps in their synthesis. **b**, A ruthenium-catalysed reaction was used to construct the B ring. **c**, **d**, The C ring was made in two steps: a palladium-catalysed reaction (**c**) followed by a gold-catalysed reaction (**d**). All three reactions are perfectly atom economical — all of the atoms in the starting materials are present in the products. Me represents a methyl group; R^1 to R^5 represent fragments of bryostatin (or of synthetic precursors to bryostatin).

The authors' reaction could, in principle, yield two products, either the desired C ring that contains six atoms, or a smaller ring made up of five atoms. Indeed, when the authors tried a palladium catalyst for the reaction, a mixture of the two rings formed. But the gold catalyst provided the six-membered ring alone.

Trost and Dong's synthesis[2] of bryostatin 16, by far the shortest to date, is a remarkable achievement. Nevertheless, more atom-economical and chemoselective methods are still needed to generate complex molecular targets with minimal effort. In particular, the development of chemoselective reactions is pivotal to reducing the need for temporary protecting groups, because every protecting group used in a synthesis generally adds two steps to the route — one to put the group on, and another to take it off. By drastically reducing the total number of steps previously required to make bryostatins, Trost and Dong's work clearly demonstrates the huge difference that chemoselective, atom-economical strategies can make. ■

André B. Charette is in the Department of Chemistry, University of Montreal, Montreal, Quebec H3C 3J7, Canada.
e-mail: andre.charette@umontreal.ca

1. Banerjee, S. *et al. J. Nat. Prod.* **71,** 492–496 (2008).
2. Trost, B. M. & Dong, G. *Nature* **456,** 485–488 (2008).
3. Pettit, G. R. *J. Nat. Prod.* **59,** 812–821 (1996).
4. Hongpaisan, J. & Alkon, D. L. *Proc. Natl Acad. Sci. USA* **104,** 19571–19576 (2007).
5. Kageyama, M. *et al. J. Am. Chem. Soc.* **112,** 7407–7408 (1990).
6. Evans, D. A. *et al. J. Am. Chem. Soc.* **121,** 7540–7552 (1999).
7. Ohmori, K. *et al. Angew. Chem. Int. Edn* **39,** 2290–2294 (2000).
8. Trost, B. M. *et al. Org. Lett.* **7,** 4761–4764 (2005).
9. Trost, B. M. & Frontier, A. J. *J. Am. Chem. Soc.* **122,** 11727–11728 (2000).
10. Gorin, D. J. & Toste, F. D. *Nature* **446,** 395–403 (2007).

MOLECULAR BIOLOGY

The Bloom's complex mousetrap

Robert M. Brosh Jr

Genomic instability often underlies cancer. Analyses of proteins implicated in a cancer-predisposing condition called Bloom's syndrome illustrate the intricacies of protein interactions that ensure genomic stability.

Bloom's syndrome, which is characterized by severe growth retardation, immunodeficiency, anaemia, reduced fertility and predisposition to cancer[1], is caused by mutations in the gene *BLM*. At the cellular level, the hallmark of this genetic disorder is a high rate of sister-chromatid exchange — the swapping of homologous stretches of DNA between a chromosome and its identical copy generated during DNA replication[2]. Understanding how mutations in *BLM* lead to this chromosomal abnormality has been of considerable interest to both scientists and clinicians. So the latest clue to solving the mystery of Bloom's syndrome, which Xu *et al.*[3] and Singh *et al.*[4] report in *Genes & Development*, is a welcome advance.

The product of *BLM*, the enzyme BLM helicase, functions as part of a protein complex of the same name, which is involved in both suppression of sister-chromatid exchange and maintenance of genomic stability. Besides this helicase, which unwinds complementary DNA double strands, this complex was thought to contain three other protein components: a topoisomerase (Topo 3α) enzyme, which unknots the two DNA strands by introducing transient nicks; RPA, which binds tightly to single-stranded DNA; and RMI1, which binds directly to BLM and Topo 3α, and perhaps less tightly to other factors.

Through its combined nicking and unwinding activities, the BLM complex catalyses the splitting (resolution) of a double-cross-shaped DNA structure called a double Holliday junction that arises from reciprocal exchanges of single strands of DNA between homologous double-stranded sequences during the process of homologous recombination (Fig. 1, overleaf). Thus, the BLM complex prevents the formation of hybrid recombinant DNA molecules called crossovers that could lead to a phenomenon known as loss of heterozygosity, which significantly contributes to cancer.

Xu *et al.*[3] and Singh *et al.*[4] independently identify a fifth component of the BLM complex: a small protein designated RMI2. Strikingly, both teams find that RMI2 is required to maintain the stability of the BLM complex, and that its deficiency in vertebrate cells results in chromosomal instability. RMI2 and BLM seem to function in the same pathway to suppress sister-chromatid exchange[3]. Moreover, RMI2-depleted cells are sensitive to DNA damage that stalls the process of replication, and this protein is essential for efficient targeting of BLM to chromatin (complexes of DNA and histone proteins) and to nuclear foci during replicational stress[4].

How does RMI2 stabilize and orchestrate the activity of the BLM complex? The answer is not simple and awaits further investigation. For starters, the current studies[3,4] suggest that RMI2 functions by interacting directly with RMI1, which, in turn, binds to BLM and Topo 3α. It is possible, however, that — in addition to the members of the BLM complex — RMI2 also binds less tightly to other protein complexes in the cell that survey and repair the genome, or to peculiar DNA configurations that arise at stalled or converging replication forks (structures formed by the separation of two complementary DNA strands during replication).
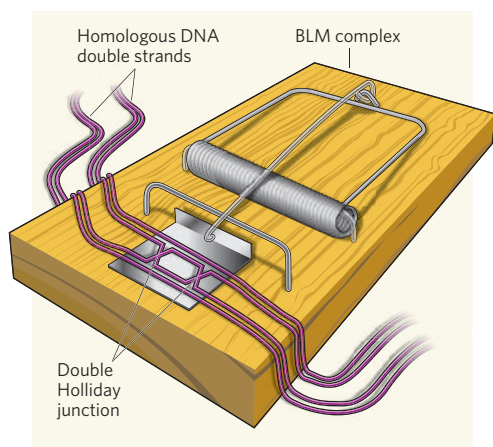
Figure 1 | Double Holliday junctions caught in the mousetrap. The BLM protein complex consists of several components, much like a mousetrap. With all the parts properly assembled, the mousetrap will operate efficiently and catch the mouse. In this case, a DNA structure called a double Holliday junction is caught in the BLM complex. Xu et al.[3] and Singh et al.[4] report the discovery of a component of this complex, RMI2, which stabilizes and orchestrates the action of the BLM complex, ensuring resolution of the double Holliday junction, and so promoting chromosomal stability.

As for the significance of RMI2 to the BLM complex, for analogy let's imagine a mousetrap. It contains several components, including a spring, a platform, a hammer, a hold-down bar and a catch[5]. Omit certain components of the trap, and the device may still operate, albeit less efficiently. With all of the components in place — including those with primarily structural roles such as the hold-down bar and the platform — the trap is most likely to catch the mouse. Returning to the BLM complex: through its interaction with RMI1, RMI2 allows the 'BLM–Topo-3α device' to assume optimal stability and configuration so that it can efficiently catalyse the splitting of the double Holliday junction, and so prevent the escape of deleterious DNA structures that would lead to crossovers (Fig. 1). RMI2 therefore seems to have an integral structural role in the BLM–Topo-3α device by orchestrating its action.

What are the implications of these observations? Mutations in the gene encoding RMI2 are likely to occur in hereditary diseases characterized by chromosomal instability and cancer. It is provocative, therefore, that there is a connection between the BLM complex and proteins mutated in another genetic disorder called Fanconi's anaemia, which also carries a high risk of cancer.

The core protein complex affected in Fanconi's anaemia (FA) and the BLM complex associate together in a supercomplex known as BRAFT (BLM, RPA, FA and Topo 3α)[6]. Singh and colleagues[4] identify FA proteins in complexes containing RMI2 and BLM, consistent with previous findings that BLM and its associated factors collaborate with FA proteins in response

to replicational stress[7,8]. So an emerging theme from these studies is that a complex network of proteins that work through overlapping and interacting pathways confers genomic integrity. To understand the functional mechanism of these protein complexes, every 'mousetrap device' must be dissected one by one. Understanding how the trap catches the mouse will allow us to grasp the consequences of chromosomal instability associated with cancer and other diseases. ■

Robert M. Brosh Jr is in the Laboratory of Molecular Gerontology, National Institute

on Aging, National Institutes of Health, NIH Biomedical Research Center, Baltimore, Maryland 21224, USA.
e-mail: broshr@mail.nih.gov

1. Ellis, N. A. et al. Cell **83,** 655–666 (1995).
2. Hanada, K. & Hickson, I. D. Cell. Mol. Life Sci. **64,** 2306–2322 (2007).
3. Xu, D. et al. Genes Dev. **22,** 2843–2855 (2008).
4. Singh, T. R. et al. Genes Dev. **22,** 2856–2868 (2008).
5. http://udel.edu/~mcdonald/mousetrap.html
6. Wang, W. Nature Rev. Genet. **8,** 735–748 (2007).
7. Hirano, S. et al. EMBO J. **24,** 418–427 (2005).
8. Pichierri, P., Franchitto, A. & Rosselli, F. EMBO J. **23,** 3154–3163 (2004).

MICROSCOPY

A terahertz nanoscope

Paul Planken

Replacing the 'micro' in microscope with 'nano', and using invisible light instead of visible, won't give scientists an instrument that can image nanostructures — unless they first beat the system's diffraction limit.

It requires a bit of imagination to think about using invisible light to look at an object of nanometre dimensions. Yet this is precisely what Huber and co-workers have done in a study described in Nano Letters[1].

Ordinary microscopes cannot see things smaller than half the size of the characteristic wavelengths present in visible light, typically a few hundred nanometres. This limitation, called the diffraction limit, becomes a great obstacle to imaging microscopically small objects with

Figure 1 | Sharp eyes of the terahertz nanoscope. Huber et al.[1] image a section through a semiconductor device comprising multiple nanoscale transistor structures. **a,** Terahertz image obtained using light with a wavelength of 118 μm. **b,** Infrared image obtained using light with a wavelength of 11 μm. Seven transistors (peaks) are clearly identified in both images. However, below the transistors, the terahertz image shows bands that correspond to the different mobile-electron densities (n) present in the device. These bands, which are absent in the infrared image, show the gain in sensitivity obtained at terahertz frequencies. (Modified from ref. 1.)

invisible, far-infrared light (also called terahertz light; 1 THz is 10^{12} Hz), because the light can have wavelengths of hundreds of micrometres — a little bigger than the thickness of a human hair. However, using techniques that were originally developed at near- and mid-infrared wavelengths to circumvent this diffraction limit, the authors have demonstrated conclusively that, even at these long wavelengths, the properties of materials can be probed on a nanometre scale. With their technique, which they call terahertz near-field nanoscopy, Huber et al.[1] have measured the concentrations of mobile electrons in a nanoscale transistor, resolving details as small as 40 nm, which is around 3,000 times smaller than the wavelength of the light used in their experiment.

Finding a way around the diffraction limit has kept scientists busy for the past few decades. To understand what the diffraction limit entails, imagine an object suspended in the air, and illuminated from behind by a light source. The presence of the object is betrayed by the shadow it casts on a wall some distance away, in the so-called far-field. It's tempting to think that if the object shrinks, the shadow on the wall will shrink in proportion to the cross-sectional area of the object. However, when the object becomes smaller than half the wavelength of the light, the shadow on the wall, in fact, disappears completely. This can be understood by remembering that light can be described as a set of waves propagating in free space. Whereas the waves are mostly blocked by large objects, they can easily bend around small ones and continue to propagate almost as if the object weren't there. This doesn't mean that waves are completely unaffected by a small object. They do, after all, have to bend around it and are thus strongly perturbed
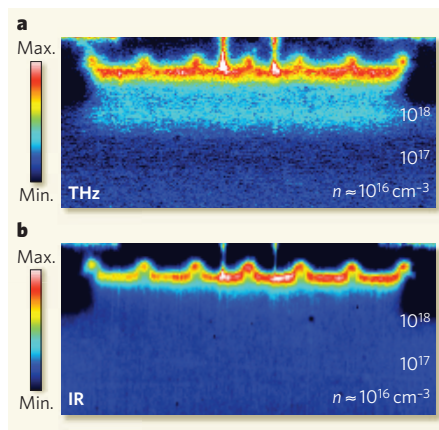
in its vicinity, just not far away from it.

In general, therefore, to see things that are smaller than half a wavelength one has to measure the light really close to the sample — that is, in the near-field region — using a very small detector. An alternative is to detect the light in the far-field while the light source, which in that case must be extremely small, is scanned very close to the sample. The latter is the approach taken by Huber *et al.* to break the diffraction limit at terahertz frequencies.

Their light source is a tiny metal tip, which tapers down to a diameter of about 30 nm, and is illuminated with a terahertz laser. This metal tip, acting as a lightning rod, collects some of the incident light and scatters it in all directions. A large portion of the scattered light is collected and measured with a sensitive detector. With some clever modulation and measurement tricks, the authors ensure that they observe light emitted only by the very end of the tip, which can thus effectively be viewed as a terahertz light source of nanometre dimensions. As it turns out, the amount of terahertz light scattered towards the detector by this source is strongly influenced by the physical properties of the sample underneath it. By scanning the tip across the sample, these physical properties, such as the concentration of mobile electrons, can be measured in unprecedented spatial detail.

Similar experiments at mid-infrared[2] and microwave[3] frequencies have been performed recently — indeed, every frequency region has its own appeal. With visible light, one can probe electronic transitions in atoms and molecules; with mid-infrared light, one can observe absorption by molecular vibrations and high-frequency lattice vibrations. So what is the added value of doing experiments at terahertz frequencies? Well, as Huber *et al.* demonstrate[1], for the typical concentrations of mobile electrons in semiconductor devices, the near-field optical contrast is actually largest in the terahertz frequency domain (Fig. 1), allowing the authors to probe as few as 100 electrons. Needless to say, this technique will find applications in anything that shows a significant response to terahertz waves, such as semiconductors, superconductors and perhaps even the low-frequency vibrations of biological molecules. With further improvements, it might even be used to characterize single molecules or electrons.

Huber and colleagues' experiment is performed at a single terahertz frequency. This is a limitation, not of the method, but rather of the terahertz source used to illuminate the tip. It's more than likely that new terahertz sources, such as the terahertz quantum-cascade laser[4] or broadband terahertz emitters[5], which can deliver radiation in a range of terahertz frequencies, will at some point be used for terahertz 'nano-spectroscopy'.

The smallest feature that can be seen with terahertz near-field nanoscopy is currently determined by the dimensions of the tip apex. However, as tips become smaller, the non-zero penetration depth of terahertz light into the metal of which the tip is made will render the tip partially transparent at its thinnest end. It will be interesting to observe how this will affect the ultimate spatial resolution of the 'terahertz near-field nanoscope'. ■

Paul Planken is in the Department of Imaging Science and Technology, Faculty of Applied Sciences, University of Technology Delft, Lorentzweg 1, Delft 2628 CJ, the Netherlands.
e-mail: p.c.m.planken@tudelft.nl

1. Huber, A. J., Keilmann, F., Wittborn, J., Aizpurua, J. & Hillenbrand, R. *Nano Lett.* **8,** 3766–3770 (2008).
2. Brehm, M. *et al. Nano Lett.* **6,** 1307–1310 (2006).
3. Ma, H. & Levy, J. *Nano Lett.* **6,** 341–344 (2006).
4. Köhler, R. *et al. Nature* **417,** 156–159 (2002).
5. van Exter, M. & Grischkowsky, D. R. *IEEE Trans. Microwave Theory Tech.* **38,** 1684–1691 (1990).

PLANT BIOLOGY

Gibberellins close the lid

Peter Hedden

Gibberellins regulate many aspects of plant growth and development. Crystal structures of their receptors provide a view in unprecedented detail of how these hormones operate at the molecular level.

We owe a lot to gibberellins. The green revolution depended on the introduction of semi-dwarfing genes that impair the biosynthesis or signalling systems of these naturally occurring plant hormones, and they or their biosynthetic inhibitors continue to be widely used to regulate crop growth. A clear understanding of how they function has emerged only recently, a notable advance being the identification[1] of a gibberellin receptor in 2005. On pages 459 and 520 of this issue, Murase *et al.*[2] and Shimada *et al.*[3] take matters further. They describe the crystal structure of receptors from two plant species, providing deeper insight into how gibberellins are perceived by plant cells.

Gibberellins (GAs) promote plant growth and developmental processes, such as seed germination and flower induction. Their action allows plants to respond to changes in their environment. At the molecular level, they stimulate the destruction of growth-repressing proteins, known as DELLA proteins[4], that bind to transcription factors and so prevent them from functioning[5,6]. The degradation of DELLA proteins requires that they are first tagged by the addition of ubiquitin molecules in a process catalysed by a protein complex known as an SCF E3 ubiquitin ligase[4]. The ubiquitinated DELLA protein is then
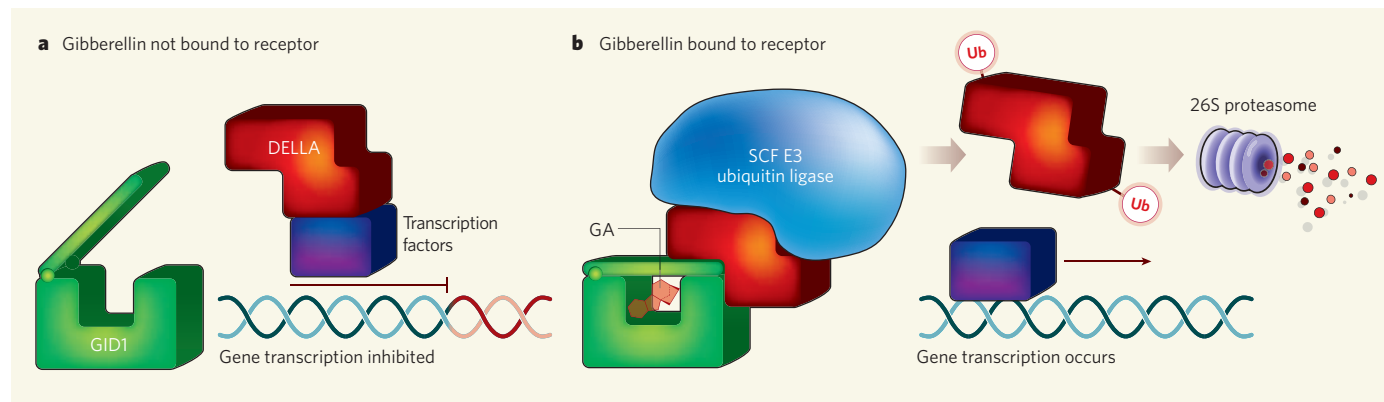

Figure 1 | Gibberellin action. a, When gibberellin (GA) concentrations are low, so that the GID1 receptor is in the unbound state, DELLA proteins inhibit gene expression by interacting with transcription factors. **b,** At higher GA concentrations, binding of GA to GID1 results in a change in protein conformation such that GID1 interacts with DELLA, promoting its association with an SCF E3 ubiquitin ligase. The resulting ubiquitination (Ub) and destruction of DELLA by the 26S proteasome frees the transcription factors to promote gene expression.
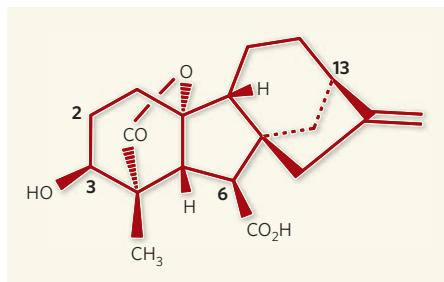
Figure 2 | Gibberellin structure. Gibberellins are diterpenoid carboxylic acids, and come in many forms. This is GA$_4$, with the carbon atoms numbered according to convention and as referred to in the text. The two new papers[2,3] show that in both GA$_3$ and GA$_4$ the carboxylic group on C6 and the hydroxyl group on C3, which are essential for biological activity, promote binding to the receptors. A hydroxyl group on C13, which is present in GA$_3$ but absent in GA$_4$, neither promotes nor hinders binding, whereas a hydroxyl group on C2 seriously reduces binding affinity. Hydroxylation on C2 is an important mechanism in higher plants for deactivating GAs.

recognized and destroyed by another protein complex, the 26S proteasome (Fig. 1).

The discovery of GA receptors has helped clarify how these hormones initiate the process of ubiquitination and degradation: in the presence of GA, the receptor, known as GID1, binds to DELLA proteins[1,7] and promotes their association with a component (the F-box) of the SCF E3 ubiquitin ligase[8]. Murase et al.[2] now present the structure of a complex comprising a GID1 receptor from the model plant *Arabidopsis thaliana* with GA and part of the DELLA protein that interacts with the receptor. Shimada et al.[3] describe the structure of the rice GID1–GA complex. In neither case was it possible to determine the structure of the receptor in isolation.

GID1 proteins resemble esterase enzymes such as the hormone-sensitive lipases that break down fat in animals. Although GID1 proteins do not function as esterases owing to a change in a critical amino acid, they have close structural similarity to these enzymes, being globular proteins containing a pocket for the substrate. The GA molecule contains four carbon rings that give it a rigid structure (Fig. 2). It is anchored by its carboxylic acid group to the bottom of the receptor pocket, such that its non-polar surface opposite the carboxylic acid group is held at the opening of the pocket. Uniquely, GID1 contains a loose strand at its amino-terminal end that interacts with the surface of the bound GA, so covering the pocket like a lid (Fig. 1).

Murase et al.[2] show that the DELLA protein interacts with the upper surface of the lid, and they speculate that this interaction may cause a change in the shape of the DELLA protein that allows it to associate with the ubiquitin ligase. Thus, GA functions as an allosteric activator of GID1, causing structural changes that allow the receptor to associate with DELLA

proteins, but it does not interact directly with DELLAs itself. The action of GA differs from that of auxin, another plant hormone, which also functions by inducing ubiquitination and degradation of transcriptional regulators known as AUX/IAAs. Auxin, however, associates directly with the F-box of the ubiquitin ligase, acting to promote its interaction with AUX/IAA without changing the structure of either protein or requiring the involvement of a third party[9].

Both papers show similar interactions of the receptors with two different GA molecules, GA$_4$ and GA$_3$. These molecules share features, including the carboxylic acid group on carbon atom 6 (C6) and a hydroxyl group on C3 (Fig. 2), that are essential for biological activity and, through interaction with polar amino-acid residues, enhance binding of the GA to the receptor. However, GA$_3$ also contains a hydroxyl group on C13, which contributes little to the binding affinity. Although most plant species predominantly use 3,13-dihydroxylated GAs, the function of the 13-hydroxyl group remains unclear. Its purpose may be to increase the solubility of the molecule and so improve mobility between cells.

The structures determined for GID1–GA indicate that a hydroxyl group on C2, which abolishes growth-promoting activity, would introduce unfavourable steric interactions with the receptor and seriously reduce binding affinity. Hydroxylation on C2 is an important mechanism in higher plants for deactivating GAs, but does not apparently occur in the more primitive club moss *Selaginella moellendorffii*[10]. In an extension of their study, Shimada et al.[3] replaced selected amino acids in the rice

GID1 protein with the corresponding amino acids in *Selaginella* GID1 and found that, in some cases, the mutated protein had lower affinity for the biologically active GA$_4$, but was more accommodating of its 2-hydroxy derivative. They propose that the receptor evolved from a hormone-sensitive lipase through loss of its catalytic activity and gradual refinement of the substrate pocket to increase affinity and specificity for GA. In higher plants, precise regulation of GA concentration is essential, and the receptor must discriminate between the active hormone and its many structurally similar biosynthetic precursors and deactivation products.

The work by Murase et al.[2] and Shimada et al.[3] has practical as well as intellectual implications, in that knowledge of the detailed structure of the receptor could help in designing more effective and potentially cheaper GA-like growth regulators for agriculture. On the intellectual front, the next challenge will be to determine how GID1–GA seals the fate of DELLAs by promoting association with the ubiquitin ligase. ■

Peter Hedden is at Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.
e-mail: peter.hedden@bbsrc.ac.uk

1. Ueguchi-Tanaka, M. et al. Nature **437**, 693–698 (2005).
2. Murase, K., Hirano, Y., Sun, T.-p. & Hakoshima, T. Nature **456**, 459–463 (2008).
3. Shimada, A. et al. Nature **456**, 520–523 (2008).
4. Ueguchi-Tanaka, M., Nakajima, M., Motoyuki, A. & Matsuoka, M. Annu. Rev. Plant Biol. **58**, 183–198 (2007).
5. Feng, S. et al. Nature **451**, 475–479 (2008).
6. de Lucas, M. et al. Nature **451**, 480–484 (2008).
7. Nakajima, M. et al. Plant J. **46**, 880–889 (2006).
8. Griffiths, J. et al. Plant Cell **18**, 3399–3414 (2006).
9. Tan, X. et al. Nature **446**, 640–645 (2007).
10. Hirano, K. et al. Plant Cell **19**, 3058–3079 (2007).

EARTH SCIENCE

# On the evolution of minerals

Minik T. Rosing

The variety of mineral species has increased since the birth of the Solar System and the development of terrestrial planets. A refreshing view likens the steady rise in mineral diversity to biological evolution.

The 94 naturally occurring chemical elements can combine into an endless number of compounds. In nature, however, such compounds come as a few thousand mineral species[1], each with its own eclectic selection of atoms systematically organized in crystal lattices. Geologists have long used the concepts of 'primitive' and 'evolved' rocks to describe the observation that planets such as Earth started out as more or less homogeneous systems, which over time segregated into chemically distinct reservoirs — on Earth, the continents, the oceans and the atmosphere, for example. The rocks that differ most markedly from the initial average planetary composition are 'evolved', and those

that never travelled far from their starting composition are 'primitive'. The planets were constructed from building blocks similar to the meteorites that fall to Earth from time to time. The meteorites, in turn, are composed of grains that condensed out of the solar nebula or formed from melt droplets that circled the early Sun. The transition from primary dust to full-scale worlds was followed by a steady increase in mineral species diversity.
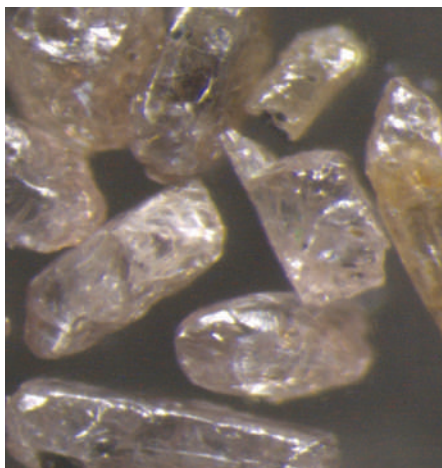
Writing in *American Mineralogist*, Hazen et al.[2] cast the progressive diversification of the mineral inventory of the Solar System and Earth (and presumably of the terrestrial planets in general) in terms of mineral evolution. They

Box 1 | A mineral selection
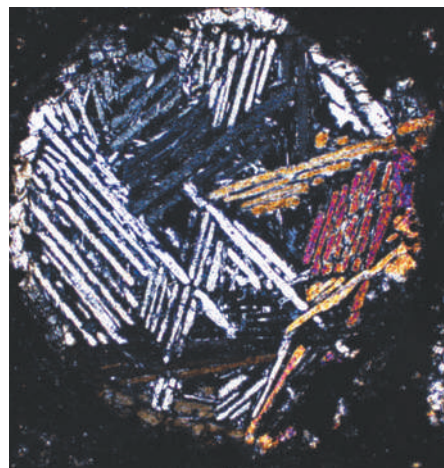
Shown here are just three examples from the wide world of mineralogy — from left to right, olivine, zircon and aragonite.

● Olivine — a magnesium–iron silicate — was one of the first minerals to form in the solar nebula, and is abundant in Earth's mantle and crust. This photomicrograph shows olivine crystals from the Allende meteorite; the crystals formed during quenching of a millimetre-sized droplet that originally circled the newborn Sun. The colours are not true, but result from optical interference seen under polarized light in the microscope.

● Zircon is especially robust and is the oldest mineral found in Earth's crust. It consists of zirconium silicate, and typically forms small, clear prisms such as these 0.1-millimetre-long crystals from Greenland. When zircons crystallize during the cooling of granite magma, they incorporate small amounts of uranium. Over time, some of the radioactive uranium decays to lead, which remains locked in the crystal. By comparing the amount of uranium remaining to the amount of new lead formed, zircon crystals can be dated with great precision and accuracy, making zircon one of the most valuable minerals for geochronology.

● The mother-of-pearl lining in the shells of some molluscs — such as these virgin paua — is a highly systematic arrangement of aragonite crystals, the size and shape of which are controlled by the animal that forms the shell. Aragonite is one of the calcium carbonate minerals often grown by organisms. Another is calcite, which forms the hard parts of, for instance, echinoderms, and the shells of several types of plankton. These shells eventually become limestone — a type of rock that constitutes some 10% of sedimentary deposits. **M.T.R.**

review the extensive literature on the mineralogy of geological materials formed through Earth's history. That history starts with cosmic dust and meteorites, and with the earliest rocks that formed from magmas. Then came rocks formed by metamorphic reactions in the heat of the deep crust, and rocks that formed at low temperatures in Earth's surface environments, and finally those minerals that are produced through the activities of living organisms. This analysis allows the division of Solar System history into three eras and ten 'mineral evolution stages', each being characterized by the emergence of domains with new physico-chemical conditions that are expressed in the form of progressively more diverse mineral assemblages.

All of the processes that combine the chemical elements into minerals are governed by a simple set of thermodynamic principles. New minerals form if energy can be released by recombination of their constituent elements from previous hosts. Minerals can only accommodate elements with specific chemical characteristics in their crystal lattices. The chemical dissimilarity between the elements becomes less pronounced with increasing temperature, so the early hot Earth had only a few minerals such as olivine (Box 1), pyroxene and feldspar. As the planet cooled, the elements found homes in more and more 'picky' mineral structures, such as clay minerals and the zeolites we use in washing powder. As Hazen *et al.* suggest, this cause of mineral species diversification must be a general feature of Earth-like planets.

On Earth itself, the emergence of life gave rise to new opportunities in the mineral world. Living organisms can construct chemical compounds that are not favoured by thermodynamic equilibrium in the ambient environment and that would not have formed spontaneously without the intervention of life. These compounds — for example aragonite in some animal skeletons — provide advantages for an organism that make it worthwhile to expend energy acquired from other life processes to synthesize them. The intervention of life in Earth's geochemical cycles therefore greatly expanded the range of mineral species in Earth's surface environments, and added minerals such as the sulphate mineral gypsum that we use in drywall panels and the haematite used in red paint. This process coupled mineral evolution to biological evolution in that some of the diversification of mineral species was caused by diversification of biological species.

It may seem problematic to use the concept of evolution in connection with the non-living world of minerals. Hazen *et al.* discuss this issue at some length and show that it is semantically and etymologically permissible to use 'evolution' to describe the change in any parameter as a function of time. This may cause unnecessary animosity towards the concept that they propose, however, which could have been avoided by using a more neutral term such as diversification or differentiation. On the other hand, the scene for this type of analogy between the living and non-living world was set many years ago by N. L. Bowen in his book *The Evolution of the Igneous Rocks*[3].

Hazen and colleagues[2] provide an excellent review of the mineral inventory of Earth and the Solar System. In addition, they provide a good sense of the evolution of the environments that permitted new minerals to form. However, their paper also includes a discussion on whether minerals can become extinct: they suggest that, as planets evolve, some minerals that were present in the early stages are succeeded by newer ones, never to recur on the scene, and that this is akin to biological extinction. But the concept of mineral extinction is unsatisfactory because, unlike living species, mineral species do not depend on the transfer of information. As a result, identical minerals will emerge on the mineralogical scene repeatedly, as long as the physical and chemical boundary conditions can be re-established, so mineral species never become universally extinct.

Finally, the authors propose that the concept of mineral evolution should be used as a framework for teaching mineralogy rather than the classic, crystallography-based system that currently forms the basis of most mineralogy instruction. I find this suggestion undesirable: the stringent systematic features of crystallography and mineral chemistry are characteristics of the physical world that help separate life from its environment.

So I have reservations about some of the ideas associated with the principle of mineral

evolution. But otherwise I think it is an exciting concept that will do much to stimulate debate and enliven thinking in the usually staid field of mineralogy.

Minik T. Rosing is at the Nordic Center for Earth Evolution, Natural History Museum of Denmark, University of Copenhagen,

DK 1350 Copenhagen, Denmark.
e-mail: minik@snm.ku.dk

1. Johnsen, O. *Minerals of the World* (Princeton Univ. Press, 2002).
2. Hazen, R. M. et al. *Am. Mineral.* **93,** 1693–1720 (2008).
3. Bowen, N. L. *The Evolution of the Igneous Rocks* (Princeton Univ. Press, 1928).

PHOTONICS

Nanomechanics gets the shakes

Tobias J. Kippenberg

Photonic circuits can allow light to be tightly confined on a chip. A clever experiment reveals how this process can be exploited to create optical forces that drive a nanoscale mechanical oscillator.

In the 1970s, Arthur Ashkin at Bell Laboratories showed[1] that the force that light exerts on an object — the radiation pressure force — can be used to trap particles. This discovery has led to a revolution in the use of lasers to trap objects ranging from atoms to biological cells. With 'optical tweezers', for example, a laser beam is focused tightly on the surface of an object, and a spatial variation in the optical intensity creates a gradient force that causes the object to be attracted to the centre of the beam, where the intensity is highest. On page 480 of this issue, Li et al.[2] describe how they have used nanoscale photonic waveguides on a silicon chip to exert optical gradient forces on a nanomechanical oscillator — a first step towards unifying nanophotonics and nanomechanics.

Observing how the radiation pressure force can affect a mechanical oscillator has been an experimental challenge for decades, although the use of nano- and micromechanical elements has allowed this weak force to be observed. For example, by using cavity optomechanical systems, in which photons are tightly coupled to the motion of an oscillator, both amplification and 'cooling' (the opposite of amplification) of the mechanical motion have been achieved[3]. But so far, researchers have observed only the scattering force that results when momentum is transferred as the light is reflected directly, and this force is parallel to the direction of light propagation. Gradient forces, in contrast, are transverse to the light beam. It has been noted[4] that gradient forces might occur in integrated photonic circuits, in which waveguides of a scale similar to the light's wavelength tightly confine the light. But these forces, which could be applied on a chip, haven't yet been observed.

Li et al.[2] use a clever design in which a single device constitutes both an optical waveguide and a nanomechanical resonator. Integration of the resonator is seamlessly accomplished by sending light through a nanophotonic waveguide, a portion of which is suspended to form a nanomechanical beam. This 'optomechanical' waveguide thereby serves both to confine light and to act as a mechanically compliant element that can vibrate at its fundamental frequency — 10 megahertz in the present study.

Crucial requirements for observing gradient forces are that the waveguides are so small that part of the optical field is evanescent (decaying exponentially with distance) and extends outside the waveguide into the substrate. The latter requirement is achieved by suspending the nanophotonic waveguide very close to a non-conducting (dielectric) substrate. In Li and colleagues' experiment, the gap between the waveguide and the (oxide) substrate is only 300 nanometres (about a fifth of the wavelength of the light used). A periodically varying gradient force can then be obtained by subjecting the light beam to a modulated power. Tuning this modulation to the resonant frequency of the nanomechanical resonator drives the oscillation of the nanomechanical beam.

Driving mechanical beams is not a new idea. What is new is exciting their motion using optical gradient forces, as this should allow a much larger bandwidth than other methods, possibly extending into the microwave regime[5,6]. Moreover, the entire actuation method can take place 'on a chip' and does not require external fields as do other methods of actuation, such as magnetomotive forces.

But how can we be sure that the gradient force really is the mechanism responsible for the oscillations? Detecting radiation pressures has long eluded researchers. One other factor known to trigger such oscillations[7,8] is the thermal effect resulting from thermal expansion, or heating of residual background gas molecules; these can mimic the effects of radiation pressure on the nanomechanical beam. A light mill, which can be bought from any toy store, is an example of a device that rotates through thermal effects rather than through radiation pressure[9].

Li et al.[2] set out to quantify thermal effects on the nanomechanical beam. Their experiments, performed in a high vacuum, reveal counterintuitively that, despite the nanoscale dimension of the waveguide, thermal gradients build up over a time period that is much longer than the period of the mechanical oscillator. Therefore, by using a sufficiently fast modulated light beam, the authors could be sure that most of the resulting periodically varying force was due to the optical field gradient rather than to thermal effects. At gigahertz frequencies, thermal effects could in fact be eliminated entirely.

The observation of gradient optical forces on a silicon chip has wide-ranging implications for the field of nanophotonics. Li and colleagues' work demonstrates that these forces can be exploited in nanophotonic circuits to trigger nanomechanical oscillations. But the principle of their experiment points to more general ways of incorporating nanomechanical and nanophotonic elements on a silicon chip. Silicon's useful mechanical properties (such as its low mechanical dissipation) have already been used to create the most sensitive mechanical beams to date[10]. But Li and colleagues' work points to a whole new class of photonic functionality — the use of gradient forces — in optically driven mechanical oscillators, photonic mixers[11] and narrow-band radio-frequency filters. As such, gradient optical forces might become a new element in 'silicon photonics'.

Probably the greatest remaining challenge is extending the method to higher resonant frequencies, which will require an increase in the optical power needed to produce the mechanical oscillations. In the authors' experiment this power is already quite high — on the order of tens of milliwatts. One solution would be to reduce the gap between the waveguide and the substrate. Alternatively, the use of cavity nano-optomechanical systems, such as integrated silicon microresonators, would not only decrease the power required by orders of magnitude, but would also provide a route to drive the mechanical oscillator without time-varying optical fields[3].

Tobias J. Kippenberg is at the Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, and is partially affiliated with the Max-Planck-Institut für Quantenoptik, Garching, Germany.
e-mail: tobias.kippenberg@epfl.ch

1. Ashkin, A. *Phys. Rev. Lett.* **24,** 156–159 (1970).
2. Li, M. et al. *Nature* **456,** 480–484 (2008).
3. Kippenberg, T. J. & Vahala, K. J. *Science* **321,** 1172–1176 (2008).
4. Povinelli, M. L. et al. *Opt. Lett.* **30,** 3042–3044 (2005).
5. Huang, X. M. H., Zorman, C. A., Mehregany, M. & Roukes, M. L. *Nature* **421,** 496 (2003).
6. Carmon, T. & Vahala, K. J. *Phys. Rev. Lett.* **98,** 223902 (2007).
7. Metzger, C. H. & Karrai, K. *Nature* **432,** 1002–1005 (2004).
8. Zalalutdinov, M. et al. *Appl. Phys. Lett.* **79,** 695–697 (2001).
9. Woodruff, A. E. *Physics Teacher* **6,** 358–363 (1968).
10. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. *Nature* **430,** 329–332 (2004).
11. Hossein-Zadeh, M. & Vahala, K. J. *Phot. Tech. Lett.* **19,** 1045–1047 (2007).

Gibberellin-induced DELLA recognition by the gibberellin receptor GID1

Kohji Murase[1,2]*, Yoshinori Hirano[1]*, Tai-ping Sun[2] & Toshio Hakoshima[1]

Gibberellins control a range of growth and developmental processes in higher plants and have been widely used in the agricultural industry. By binding to a nuclear receptor, GIBBERELLIN INSENSITIVE DWARF1 (GID1), gibberellins regulate gene expression by promoting degradation of the transcriptional regulator DELLA proteins, including GIBBERELLIN INSENSITIVE (GAI). The precise manner in which GID1 discriminates and becomes activated by bioactive gibberellins for specific binding to DELLA proteins remains unclear. Here we present the crystal structure of a ternary complex of *Arabidopsis thaliana* GID1A, a bioactive gibberellin and the amino-terminal DELLA domain of GAI. In this complex, GID1A occludes gibberellin in a deep binding pocket covered by its N-terminal helical switch region, which in turn interacts with the DELLA domain containing DELLA, VHYNP and LExLE motifs. Our results establish a structural model of a plant hormone receptor that is distinct from the mechanism of the hormone perception and effector recognition of the known auxin receptors.

More than 70 years have passed since gibberellin (GA) was identified as a fungus toxin from *Gibberella fujikuroi* (Sawada), which causes the 'bakanae' (foolish seedling) disease of rice, a condition manifested by hyper-elongated stems and male sterility that markedly lowers the yield of rice crops[1]. In the 1950s, GAs were recognized as extremely important endogenous hormones in higher plants[2]. Since then, the precise mechanism by which the GA signal is perceived and interpreted by plant cells has remained a longstanding question that is central in plant biology.

Bioactive GAs are a group of diterpenoid acids that regulate diverse developmental processes throughout the life cycle of a plant[3]. Out of the 136 GAs identified from plants, fungi and bacteria (http://www.plant-hormones.info/gibberellins.htm)[4], only a few such as GA_1, GA_3, GA_4 and GA_7 function as bioactive hormones. These bioactive GAs are all hydroxylated at C3 and contain a lactone ring between C4 and C10 as well as a carboxyl group at C6 (Fig. 1a and Supplementary Fig. 1a). Addition of a hydroxyl group at the C2 position deactivates GAs. However, the lack of structural knowledge of GA receptors prevents our understanding of the precise structure–activity relationships pertaining to these molecules.

Breakthroughs in GA-signalling research have included identification of the soluble GID1 GA receptors[5,6], the transcriptional regulators DELLA proteins, which negatively regulate GA signalling[7,8], and an F-box protein[9,10]. DELLA proteins contain conserved DELLA and VHYNP sequences in the N-terminal regulatory region, which are important for GID1 binding and proteolysis of the DELLA proteins[11–15], and the conserved GRAS (for GAI, RGA and SCARECROW) domain[16] in the carboxy-terminal half.

On GA binding, the GA receptor GID1 is activated, which leads to the recognition of DELLA proteins through an unknown mechanism.
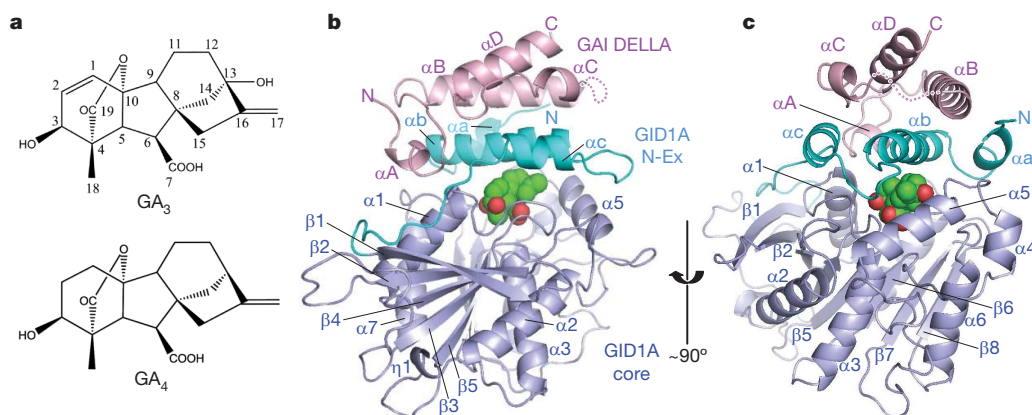
Figure 1 | Structure of the GA₃–GID1A–DELLA complex. **a**, Chemical structures and numbering of GA₃ and GA₄ that were used in our studies. **b**, Ribbon representation of the GA₃–GID1A–DELLA complex, with the GAI DELLA domain (pink), the GID1A N-terminal extension (N-Ex, cyan) and the GID1A α/β core domain (light blue). The bound GA₃ molecule is represented as a space-filling model with carbon in green and oxygen in red. **c**, As in **b**, but with a rotation of ~90° around a vertical axis. Three N-terminal extension helices αa–αc of GID1A are loosely packed against each other to form a flat cover on the pocket.

[1]Structural Biology Laboratory, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan. [2]Department of Biology, Duke University, Durham, North Carolina 27708, USA.
*These authors contributed equally to this work.

This binding enables DELLA proteins to be recruited by the F-box protein in a ubiquitin ligase E3 SKP1–CULLIN–F-box (SCF) complex for polyubiquitylation, followed by degradation by means of the 26S proteasome[9,10,13,17–19]. GA–GID1-induced DELLA degradation allows transcriptional reprogramming of GA-responsive genes[20].

A. thaliana contains 3 functionally redundant GA receptors (GID1A, GID1B and GID1C)[6,13,14,21], 5 DELLA proteins and at least 23 GA derivatives[22,23]. The DELLA proteins comprise GAI (ref. 7), REPRESSOR OF *ga1-3* (RGA)[8] and RGA-LIKE1, 2 and 3, which display partially redundant functions in modulating vegetative and reproductive growth[24–28].

Here we report on a series of crystallographic and biochemical studies of GA perception by the *A. thaliana* GA receptor GID1A and GA-induced binding of GID1A to the DELLA protein GAI. Our results reveal a structural mechanism that underlies both GA sensing and recognition of DELLA proteins by GID1, and suggest a new direction for the development of GA derivatives as new growth regulators that target crops and ornamental plants.

Overall structure of the ternary complex

The crystal contains GA$_3$-bound GID1A (residues 1–344) and the N-terminal GAI DELLA domain (residues 11–113) encompassing DELLA and VHYNP sequences, hereafter referred to as the GA$_3$–GID1A–DELLA complex. Application of GA$_3$ to *Arabidopsis* showed that it is an active GA, although it does not seem to be an endogenous hormone in this plant. GA$_4$ is the major bioactive GA in *Arabidopsis*. Thus, we determined both the GA$_3$– and the GA$_4$–GID1A–DELLA complex structures at 1.8 ångström (Å) resolution (Supplementary Table 1). The structures were essentially the same, with the averaged root mean squared deviation in all Cα atoms of 0.18 Å (Supplementary Fig. 1b). Our discussion will therefore focus on the structure of the GA$_3$–GID1A–DELLA complex followed by reference to a local structural difference in the two GAs.

The GA$_3$–GID1A–DELLA complex displays a stocky structure with a globular GID1A that is bound on one side by the GAI DELLA domain (Fig. 1b, c). GID1A is a monomeric protein and comprises one α/β core domain with the N-terminal extension that extends up the core surface towards the DELLA domain. The N-terminal extension helices are sandwiched between the DELLA domain and the GID1A core domain, providing the molecular glue between the two domains (Fig. 1b and Supplementary Fig. 1c). Our crystal structures indicate that GID1A, but not DELLA, is responsible for GA perception. The GID1A core domain possesses a central deep pocket that accommodates GA$_3$ and is covered with the N-terminal extension. This binding mode insulates the bound hormone molecule from both the solvent region and the DELLA protein: there is no direct contact between the hormone and the DELLA domain.

GID1A architecture

The GID1A core domain consists of an eight-stranded mixed β-sheet, which comprises one anti-parallel and seven parallel β-strands, and seven α-helices as well as two 3$_{10}$ helices (Fig. 1b). The β-sheet exhibits a left-handed superhelical twist. The architecture indicates that the GID1A core domain belongs to the plant carboxylesterase family of the α/β-hydrolase fold superfamily, as previously noted[29]. The GID1A structure is found to be most similar to the plant carboxylesterase from *Actinidia eriantha* (CXE1)[29], the core domain of which superimposes well with GID1A (Supplementary Fig. 2). GID1A lacks three α-helices corresponding to α3, α5 and α8 of CXE1 (Supplementary Fig. 3).

Unlike the carboxylesterases, GID1A possesses a unique N-terminal extension (1–62 residues), which is highly conserved among GID1 members, implying its specific function as the GA receptor. The extension forms three helices (αa–αc) and an extended loop (loop αc–β1; Fig. 2a, b). Three helices αa–αc are loosely packed to form a flat cover on the GA-binding pocket and bury a broad surface (2,400 Å^2) on the core domain. The interfaces comprise both polar and nonpolar contacts and contain many bridging water molecules (Fig. 2c).
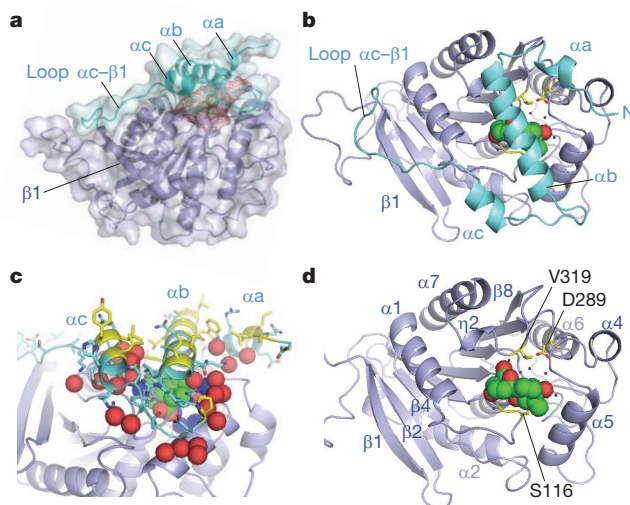
Figure 2 | Architecture of GID1A. a, The N-terminal extension (cyan) covers the GID1A pocket (a red cage). **b**, A top view of the GID1A core domain (light blue), with the N-terminal extension (cyan). **c**, The interface between the N-terminal extension helices and the GID1A core domain contains many water molecules (red spheres). Nonpolar residues projecting outside are represented as yellow stick models. **d**, As in **b**, but without the N-terminal extension. Val 319 and Asp 289 of the triad residues are represented as stick models. Asp 289 is hydrated with water molecules (blue small spheres). In **b**–**d**, the GA$_3$ molecule is represented as a space-filling model (green).

Intriguingly, the helices have nonpolar residues projected outside, which are mostly conserved in GID1s and participate in interactions with the DELLA domain, as will be described. The core domain preserves an extra helix α5, which has a role as part of the pocket wall rather than as a lid (Figs 1c and 2d).

GID1A contains two out of three conserved active residues for carboxylesterase in the deep pocket: the nucleophilic active residue Ser 191 in the GxSxG motif, and Asp 289 (Fig. 2d). The histidine residue of the Ser-His-Asp triad, however, is replaced with Val 319, thus GID1A exhibits no esterase activity[6]. GID1A contains the conserved Gly 115-Ser 116 pair, which forms the oxyanion hole with Ser 192. At the bottom of the pocket, Asp 289 is hydrated with water molecules, which fill the space at the bottom of the pocket (Fig. 2d). In total, six water molecules are sequestered in the pocket and form a hydrogen bonding network.

Gibberellin perception by GID1A

The embedded GA$_3$ molecule exhibits a large contact surface area (753 Å^2) by directing the hydrophilic carboxylate towards the bottom and the hydrophobic aliphatic rings to the entrance of the pocket. The GA$_3$ aliphatic rings contact the surrounding hydrophobic walls at the pocket entrance formed by nonpolar residues (Ile 126, Leu 323, Val 239 and Val 319; Fig. 3a). The N-terminal extension helices of GID1A project nonpolar residues (Ile 24, Phe 27 and Tyr 31 from helix αb) inside to recognize the shape of the GA fused rings by wedging their side chains into grooves formed by two of the fused rings and sandwiching the rings (Fig. 3b). It should be noted that these residues and His 119 are tightly packed with each other, so that the position and orientation of the side chains are strictly adjusted for GA ring recognition.

The active GA$_3$ molecule possesses one carboxylate, two hydroxyl and one lactone moiety (Fig. 1a). All of these polar groups participate in direct interactions with GID1A and water molecules, where a total of ten hydrogen bonds are formed including three bridging water molecules (Supplementary Fig. 4 and Supplementary Table 2). The 7-carboxylate anchors the hormone molecule to the bottom of the
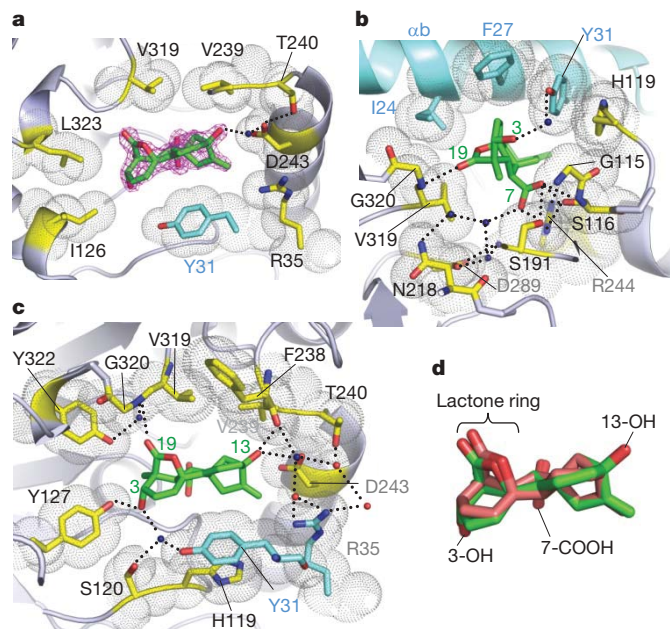
Figure 3 | Recognition of GA$_3$ by GID1A. a, A top view of the GID1A-bound GA$_3$. GID1A residues that contact GA$_3$ are highlighted in yellow (the core domain) or cyan (the N-terminal extension) and shown as stick models with van der Waals surfaces (dots) and hydrogen bonds (dotted lines). Omit electron density for GA$_3$ is shown in 3.5σ contour (red). **b,** A side view showing nonpolar contacts between N-terminal extension helix αb and the GA$_3$ aliphatic rings. **c,** A top view of the hydroxyl groups of GA$_3$. **d,** The bound GA$_3$ (green) superimposes onto GA$_4$ (pink) with a small root mean squared deviation (0.20 Å). Water molecules are shown as blue or red spheres.

pocket by forming multiple hydrogen bonds with two residues (Ser 116 and Ser 191) from the pocket floor and one water molecule (Fig. 3b). Moreover, the carboxylate negative charges are neutralized by the oxyanion hole and by forming a salt bridge with Arg 244 (from helix α5), which is conserved in GID1s but not in carboxylesterases including CXE1. The 3-hydroxyl group, which is necessary for GA activity, is hydrogen-bonded to Tyr 127 and a bridging water molecule (Fig. 3c and the GA$_4$–GID1A–DELLA complex in Supplementary Fig. 5).

The 13-hydroxyl group, which is common in bioactive GA$_1$ and GA$_3$ but absent in GA$_4$, forms weak hydrogen bonds to the main chain (Phe 238) and a bridging water molecule (Fig. 3c). The 13-hydroxyl group, however, is located close to the negatively charged Asp 243, which might account for weaker binding affinity of GA$_3$ and GA$_1$ than that of GA$_4$ (ref. 6). Despite having these differences, the bound GA$_3$ molecule is well overlapped with GA$_4$ in the GA$_4$–GID1A–DELLA complex (Fig. 3d), minimizing differences in the GID1A structure for ensuring bioactivity.

The most characterized *in vivo* deactivation of bioactive GAs relates to introduction of the 2-hydroxyl group[30], which should cause steric interference with Tyr 31 and introduce ring conformational changes unfavourable for receptor binding. Methylation of the 7-carboxylate by methyltransferases[31] reduces the affinity by two to three orders of magnitude[6]; this is due to the disruption of multiple hydrogen bonding and ionic interactions with the oxyanion hole and GID1-specific arginine as described previously. Mutation of these key residues markedly reduces the binding affinity[15].

The DELLA–GID1A interaction

The DELLA domain of GAI forms four α-helices, αA to αD (Fig. 4a). The N-terminal region forms a short helix αA followed by a reverse turn and a loop connected to helix αB. Two long helices αB and αD run in parallel polarity, with the helical axes inclined by ~20° to pack
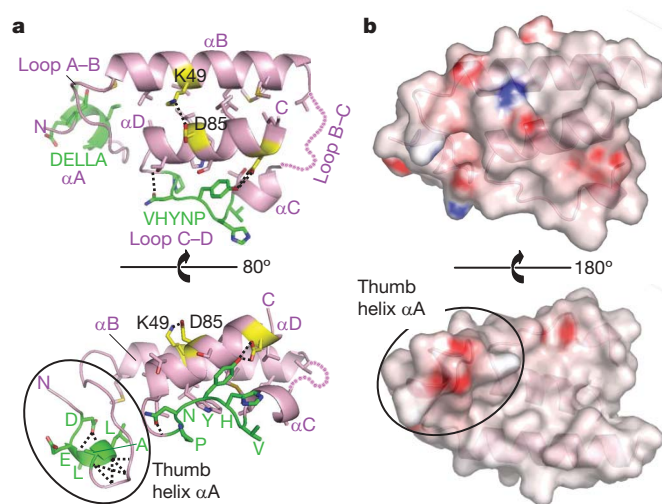
Figure 4 | Architecture of the DELLA domain of GAI. a, Top (front) and side views of the DELLA domain of GAI consists of four α-helices: αA (residues 28–31), αB (43–57), αC (69–72) and αD (80–91). Sequence motifs believed to be important for GID1 binding are highlighted in green. Intra-domain hydrogen bonds are indicated with black dotted lines. The thumb-like part containing helix αA is circled. **b,** Front and back views of the DELLA domain of GAI with electrostatic potential surface representation. The front view corresponds to the top view in **a**. Positively and negatively charged potentials are in blue and red, respectively.

the 'knobs into holes'[32] of the nonpolar aliphatic side chains and are stabilized by an inter-helical salt bridge (Lys 49 and Asp 85)[33]. The poorly conserved loop between helices αB and αC (Supplementary Fig. 6) is disordered in the current map. The flat-shaped DELLA domain resembles a palm consisting of helices αB to αD, with helix αA being off-plane like a thumb. The outside forms a negatively charged surface whereas the inside forms a nonpolar surface, which interacts with GID1A (Fig. 4b). The DELLA sequence (residues 28–32) is located within helix αA, which is N-capped by Asp 28 and C-capped by the main-chain amides of the following reverse turn (residues 32–35). The VHYNP motif is located within loop C–D, where Y and N of the motif participate in intra-domain interactions that maintain the loop conformation.

Overall, the DELLA domain buries an accessible surface area of 2,620 Å^2 on GID1A. Notably, the DELLA palm contacts the GID1A N-terminal extension helices by nonpolar interactions, whereas helix αA interacts with both the N-terminal extension and the core domain (Fig. 5a). DELLA binding stabilizes the GID1A extension helices covering the pocket that traps GA, causing an apparent enhancement of GA binding to GID1A[6]. We define three conserved motifs of the DELLA domain that are essential for direct contact with the hormone receptor.

The structure enables us to extend the DELLA motif and define a GID1 recognition sequence comprising DeLLaΦLxYxV (in which capital letters are residues that directly contact GID1, Φ represents a nonpolar residue, and lowercase letters are residues that do not participate in direct contacts, where 'e' and 'a' represent Glu and Ala residues, respectively, and 'x' represents any residue). The first aspartate residue (Asp 28) of the DELLA motif forms a hydrogen bond with Asn 19 of the GID1A N-terminal extension. The DELLA thumb helix αA is hooked onto the nonpolar crevice of GID1A formed by the N-terminal extension and the core domain (Fig. 5b). Thus, this motif should significantly contribute to the binding affinity. In fact, deletion of this motif leads to a GA-insensitive phenotype[7,11,12].

After the DELLA motif, we highlight the LExLE motif on helix αB with the consensus sequence MAxVAxxLExLExΦ; all the nonpolar residues in this motif pack against GID1A helix αb (Fig. 5c). Notably, two negatively charged glutamates of the motif stabilize binding by forming salt bridges with GID1A positively charged
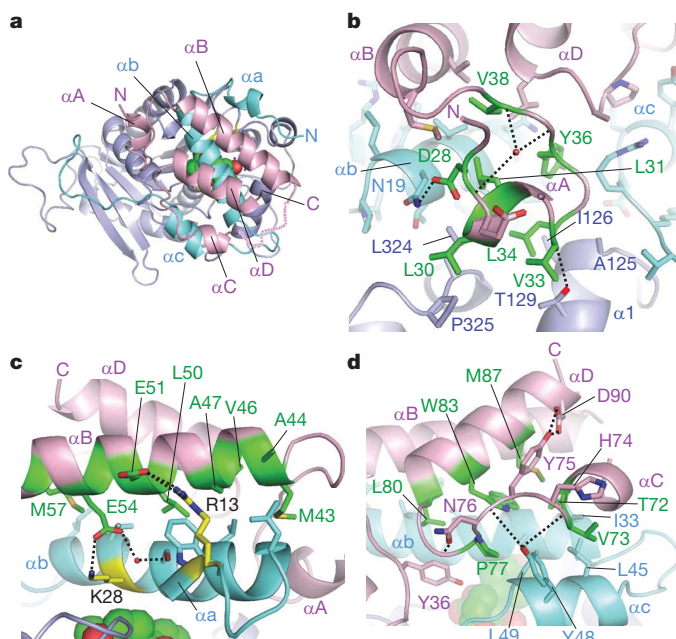
Figure 5 | Recognition of the DELLA domain by GID1A. a, A top view of the GID1A core domain (light blue) and the N-terminal extension (cyan) with the GAI DELLA domain (pink). The view corresponds to Fig. 2b, d. **b,** A close-up view of the thumb helix αA hooked onto the GID1A crevice. Residues of the DELLA motif (residues D28–V38) are highlighted in green. Hydrogen bonds are indicated with dotted lines. A bridging water molecule inside the loop is shown as a red sphere. **c,** The LExLE motif (M43–M57). **d,** The VHYNP motif (T72–M87). In **b–d,** the residues in each motif that directly contact GID1A are highlighted in green.

Figure 6 | A model of GA-regulated GID1–DELLA protein interactions. GA functions as an 'allosteric inducer' to enhance GID1–DELLA interactions. GA binding induces a GID1 conformational change in the N-terminal helical switch for DELLA binding, which promotes a conformational transition in the DELLA protein. Thus, GA-activated GID1 functions as a 'ubiquitylation chaperon' that stimulates substrate recognition by the SCF complex.

the GRAS domain, which enhances binding to the F-box protein SLY1 that targets the DELLA protein for degradation[13,18]. However, delineation of the precise nature of the intramolecular interactions between the DELLA and GRAS domains requires future structural studies.

Implication in plant and general biology

A common strategy used in the transmission of several plant hormonal signals is the removal of key negative regulators using specific SCF ubiquitin ligase in a hormone-concentration-dependent manner[34,35]. The protein under hormone-dependent degradation is often a repressor, such as with auxin, GA and jasmonate signalling pathways, or an activator, as with ethylene signalling. Auxin and jasmonate receptors are F-box proteins[36–39] and, recently, auxin was shown to serve as the 'molecular glue' that brings together a substrate protein and the F-box protein[40]. In contrast, the GA receptor is a genuine receptor protein that is activated by the allosteric effector GA to function as a 'ubiquitylation chaperon' that stimulates substrate recognition by the SCF complex. Thus, the perception mechanisms of these hormones differ. Remarkably, both mechanisms are novel in that they generate an exposed degradation signal in the protein to be degraded, compared with other mechanisms that involve phosphorylation by protein kinases, unmasking by protein dissociation or creation of a destabilized N terminus by proteolysis[41]. Thus, the structural investigation of plant hormone receptors has deepened our understanding of the degradation-signal-dependent regulation of cellular functions.

residues (Arg 13 and Lys 28). The mutation of key residues of the LExLE motifs markedly reduces binding of the DELLA domain to GA$_3$-bound GID1A (Supplementary Fig. 7).

The VHYNP motif extends to helix αD, with an extended motif TVhynPxxLxxWxxxM providing nonpolar residues that directly contact GID1A. The VHYNP motif sandwiches the aromatic ring of Tyr 48 from GID1A helix αc with Val and Pro residues and forms a hydrogen bond to the phenolic hydroxyl group (Fig. 5d). Although the HYN residues have no direct contact with GID1A, they may contribute to stabilization of helix αD and the following loop conformation.

Mechanism of GA-dependent activation

Our complex structures and the following biochemical and spectroscopic studies support the idea that bioactive GA is an allosteric inducer of GID1A, the latter adopting a compact form with the N-terminal extension being folded back to cover the GA-bound pocket. Our crystallization trials of GA-unbound GID1A were unsuccessful, probably due to the flexible nature of this protein in solution. Protease digestion experiments further revealed that GID1A is highly sensitive to proteases in the unbound state but resistant in the GA-bound state and even more resistant in the DELLA-bound state (Supplementary Fig. 8a). In addition, the helix bundle of the DELLA domain may be induced by GID1A binding. Indeed, our circular dichroism spectra suggest a random-coil conformation of the isolated DELLA domain (Supplementary Fig. 8b). Although our knowledge concerning the DELLA domain conformation of the full-length DELLA protein is limited, we speculate that the DELLA domain conformation is also unstable or metastable. On the basis of these observations and consideration, we propose that the N-terminal extension of GID1A acts as a conformational switch that senses GA. In the 'ON' state, the folded N-terminal extension creates binding surfaces for the DELLA domain and induces a coil-to-helix conformational transition of the DELLA domain (Fig. 6). The transition of the DELLA domain may induce a conformational change in

METHODS SUMMARY

also refined at 1.8 Å resolution. Pull-down binding assays were performed in a manner similar to that described previously[13].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Yabuta, T. & Sumiki, Y. On the crystal of gibberellin, a substance to promote plant growth. *J. Agric. Chem. Soc. Japan* **14,** 1526 (1938).
2. Radley, M. Occurrence of substances similar to gibberellic acid in higher plants. *Nature* **178,** 1070–1071 (1956).
3. Yamaguchi, S. Gibberellin metabolism and its regulation. *Annu. Rev. Plant Biol.* **59,** 225–251 (2008).
4. MacMillan, J. Occurrence of gibberellins in vascular plants, fungi, and bacteria. *J. Plant Growth Regul.* **20,** 387–442 (2002).
5. Ueguchi-Tanaka, M. et al. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* **437,** 693–698 (2005).
6. Nakajima, M. et al. Identification and characterization of *Arabidopsis* gibberellin receptors. *Plant J.* **46,** 880–889 (2006).
7. Peng, J. et al. The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* **11,** 3194–3205 (1997).
8. Silverstone, A. L., Ciampaglio, C. N. & Sun, T.-p. The *Arabidopsis RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* **10,** 155–169 (1998).
9. McGinnis, K. M. et al. The *Arabidopsis SLEEPY1* gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell* **15,** 1120–1130 (2003).
10. Sasaki, A. et al. Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. *Science* **299,** 1896–1898 (2003).
11. Dill, A., Jung, H.-S. & Sun, T.-p. The DELLA motif is essential for gibberellin-induced degradation of RGA. *Proc. Natl Acad. Sci. USA* **98,** 14162–14167 (2001).
12. Itoh, H., Ueguchi-Tanaka, M., Sato, Y., Ashikari, M. & Matsuoka, M. The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant Cell* **14,** 57–70 (2002).
13. Griffiths, J. et al. Genetic characterization and functional analysis of the GID1 gibberellin receptors in *Arabidopsis. Plant Cell* **18,** 3399–3414 (2006).
14. Willige, B. C. et al. The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of *Arabidopsis. Plant Cell* **19,** 1209–1220 (2007).
15. Ueguchi-Tanaka, M. et al. Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. *Plant Cell* **19,** 2140–2155 (2007).
16. Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D. & Benfey, P. N. The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the *SCARECROW-LIKE* genes. *Plant J.* **18,** 111–119 (1999).
17. Silverstone, A. L. et al. Repressing a repressor: gibberellin-induced rapid reduction of the RGA protein in *Arabidopsis. Plant Cell* **13,** 1555–1565 (2001).
18. Dill, A., Thomas, S. G., Hu, J., Steber, C. M. & Sun, T.-p. The *Arabidopsis* F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. *Plant Cell* **16,** 1392–1405 (2004).
19. Fu, X. et al. The *Arabidopsis* mutant sleepy1[gar2-1] protein promotes plant growth by increasing the affinity of the SCF[SLY1] E3 ubiquitin ligase for DELLA protein substrates. *Plant Cell* **16,** 1406–1418 (2004).
20. Zentella, R. et al. Global analysis of DELLA direct targets in early gibberellin signaling in *Arabidopsis. Plant Cell* **19,** 3037–3057 (2007).
21. Iuchi, S. et al. Multiple loss-of-function of *Arabidopsis* gibberellin receptor AtGID1s completely shuts down a gibberellin signal. *Plant J.* **50,** 958–966 (2007).
22. Talon, M., Koornneef, M. & Zeevaart, J. A. D. Endogenous gibberellins in *Arabidopsis thaliana* and possible steps blocked in the biosynthetic pathways of the semidwarf *ga4* and *ga5* mutants. *Proc. Natl Acad. Sci. USA* **87,** 7983–7987 (1990).
23. Derkx, M. P. M., Vermeer, E. & Karssen, C. M. Gibberellins in seeds of *Arabidopsis thaliana*: biological activities, identification and effects of light and chilling on endogenous levels. *Plant Growth Regul.* **15,** 223–234 (1994).
24. King, K. E., Moritz, T. & Harberd, N. P. Gibberellins are not required for normal stem growth in *Arabidopsis thaliana* in the absence of GAI and RGA. *Genetics* **159,** 767–776 (2001).
25. Dill, A. & Sun, T.-p. Synergistic derepression of gibberellin signaling by removing RGA and GAI function in *Arabidopsis thaliana. Genetics* **159,** 777–785 (2001).
26. Lee, S. et al. Gibberellin regulates *Arabidopsis* seed germination via *RGL2*, a *GAI/RGA-like* gene whose expression is up-regulated following inhibition. *Genes Dev.* **16,** 646–658 (2002).
27. Cheng, H. et al. Gibberellin regulates *Arabidopsis* floral development via suppression of DELLA protein function. *Development* **131,** 1055–1064 (2004).
28. Tyler, L. et al. DELLA proteins and gibberellin-regulated seed germination and floral development in *Arabidopsis. Plant Physiol.* **135,** 1008–1019 (2004).
29. Ileperuma, N. R. et al. High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon. *Biochemistry* **46,** 1851–1859 (2007).
30. Thomas, S. G., Phillips, A. L. & Hedden, P. Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc. Natl Acad. Sci. USA* **96,** 4698–4703 (1999).
31. Varbanova, M. et al. Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* **19,** 32–45 (2007).
32. Crick, F. H. C. The packing of α-helices: simple coiled coils. *Acta Crystallogr.* **6,** 689–697 (1953).
33. Hakoshima, T. in *Nature Encyclopedia of the Human Genome* Vol. 3, 679–683 (Nature Publishing Group, 2003).
34. McSteen, P. & Zhao, Y. Plant hormones and signaling: common themes and new developments. *Dev. Cell* **14,** 467–473 (2008).
35. Ho, M. S., Ou, C., Chan, Y.-r., Chien, C.-T. & Pi, H. The utility F-box for protein destruction. *Cell. Mol. Life Sci.* **65,** 1977–2000 (2008).
36. Dharmasiri, N., Dharmasiri, S. & Estelle, M. The F-box protein TIR1 is an auxin receptor. *Nature* **435,** 441–445 (2005).
37. Kepinski, S. & Leyser, O. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* **435,** 446–451 (2005).
38. Thines, B. et al. JAZ repressor proteins are targets of the SCF[COI1] complex during jasmonate signalling. *Nature* **448,** 661–665 (2007).
39. Chini, A. et al. The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448,** 666–671 (2007).
40. Tan, X. et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446,** 640–645 (2007).
41. Alberts, B. et al. in *Molecular Biology of the Cell* 5th edn, 395–396 (Garland Science, 2008).

Author Contributions K.M., T.-p.S and T.H. conceived and designed the project. T.-p.S. provided all cDNA and constructs as well as expertise in the GA signalling mechanism. Y.H., K.M. and T.H. were responsible for construct design for protein preparation. K.M. was responsible for subcloning and execution of protein biochemistry, crystallization and data collection, helped by Y.H. and directed by T.H. Y.H. solved and refined the complex structures. K.M. performed binding studies with GID1A and DELLA variants, as well as circular dichroism measurements, directed by Y.H. Y.H, K.M. and T.H. interpreted data and T.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Structure coordinates and structural factors are deposited in the Protein Data Bank under accession numbers 2ZSH (GA₃–GID1A–DELLA) and 2ZSI (GA₄–GID1A–DELLA). Correspondence and requests for materials should be addressed to T.H. (hakosima@bs.naist.jp).

METHODS

Protein expression and purification. A polymerase chain reaction (PCR)-amplified *GID1A* fragment (residues 1–344) was cloned into the BamHI/SacI site of pET-47b[+] (Novagen) to generate pET47GID1A2. The NdeI/AvrII fragment containing *GID1A* with an N-terminal hexahistidine tag (His-tag) and human rhinovirus 3C (HRV 3C) protease recognition site from pET47GID1A2 was subcloned into the NdeI/AvrII site of pCDFDuet-1 (Novagen), resulting in pCDFa2x. PCR-amplified *GAI* fragments coding for the DELLA domain (residues 11–113) were cloned into the BamHI/SacI site of pET-47b[+] to generate pET47GAI6. pCDFa2x and pET47GAI6 were co-transformed into *E. coli* strain BL21Star (DE3, Invitrogen).

Protein co-expression was performed at 16 °C for 24 h in Luria-Bertani medium containing 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 0.1 mM GA₃ (Tokyo Chemical Industry). Cells were collected, suspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM DTT) containing 0.1 mM GA₃, and lysed by sonication. After ultracentrifugation at 100,000*g* for 30 min, the supernatant was applied to a Ni–NTA resin (Qiagen). Proteins were eluted with lysis buffer containing 250 mM imidazole and treated with PreScission protease (GE Healthcare) for His-tag removal. Proteins were further purified by anion exchange (HiTrap Q HP, GE Healthcare) and gel-filtration chromatography (Superdex 75 pg, GE Healthcare). For preparation of the selenomethionine (SeMet)-labelled GID1A–GAI complex, pCDFa2x and pET47GAI6 plasmids were co-transformed into *E. coli* strain B834 (DE3, Novagen). Protein expression was performed in LeMaster medium containing 0.18 mg l⁻¹ SeMet[42]. The expression conditions and purification steps were the same as those used for the native protein. The protein complex, comprising the GA₃-bound GID1A complexed with the GAI DELLA domain, was stable during purification, as determined by polyacrylamide gel electrophoresis with sodium dodecyl sulphate (SDS–PAGE) and matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI-TOF MS; PerSeptive Inc.). Gel filtration suggests that the complex exists as a monomer in solution (data not shown). MALDI-TOF MS of the native protein complex gave peaks at 39,221 Da (calculated 39,228 Da) and 12,273 Da (12,272 Da) for the GID1A and the DELLA domain, respectively, and that of the SeMet complex gave 39,407 Da (39,415 Da) and 12,659 Da (12,647 Da), respectively.

Crystallization and data collection. Crystallization conditions were searched for using the sitting-drop vapour diffusion method and Hydra II-Plus-One crystallization robot (Matrix Technology) with a commercial crystallization solution kit. The best crystals were obtained from a 2:1 mixture solution comprising 50 mg ml⁻¹ of the complex solution (in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT with 0.1 mM GA₃) and reservoir solution (0.1 M Tris-HCl, pH 8.0, 0.8 M LiCl₂ and 26% polyethylene glycol 4000). The GID1A–GAI complex crystals appeared within 3–4 weeks at 20 °C. The 300–400 μm crystals were transferred into cryoprotectant solution containing mother liquor with 10% polyethylene glycol 200 and then flash-cooled in liquid nitrogen. The SeMet-labelled complex was crystallized under the same conditions as for the native complex. Diffraction tests of the crystals obtained were performed at 100 K using a Rigaku R-AXIS VII detector equipped with a Rigaku FR-E X-ray generator. For structure determination, diffraction data of native and SeMet-labelled crystals were collected at 100 K on a BL41XU beamline at the SPring-8 synchrotron facility. Data were collected at the absorption peak point. The total oscillation ranges were 360° for peak and 180° for native data sets, respectively. The diffraction data were processed using the HKL2000 program[43]. The crystal belongs to space group *P*4₃2₁2 with unit cell parameters *a* = 82.0 Å, *b* = 82.0 Å and *c* = 130.1 Å and a Matthews coefficient (*V*$_M$) of 2.1 Å³ per Da, suggesting a solvent content of 42% assuming that one complex is present in the asymmetric unit.

Structure determination and refinement. Phases were calculated by a single-wavelength anomalous dispersion method using data collected at the peak wavelength of selenium. Out of a total of 12 sites, 6 selenium atoms were located using the program BnP[44] and solvent flattening was performed with RESOLVE[45]. The built model was refined through alternating cycles using the Coot[46] and CNS programs[47]. The model has been refined to 1.8 Å resolution with *R*$_{work}$ and *R*$_{free}$ values of 20.5% and 22.9%, respectively. Peptide chains corresponding to residues 6–343 of GID1A and residues 25–91 of GAI were modelled along with a GA₃ molecule and 217 water molecules. The first five N-terminal residues and one C-terminal residue of GID1A, in addition to the three regions corresponding to residues 11–24, 61–67 (a loop between helices αB and αC (loop B–C)) and 92–113 of GAI, were omitted from the current model.

The GA₄–GID1A–DELLA complex was also prepared and crystallized in the same manner as the GA₃–GID1A–DELLA complex. The structure was determined by molecular replacement using the GA₃–GID1A–DELLA complex structure and refined at 1.8 Å resolution. The crystallographic data and refinement statistics are summarized in Supplementary Table 1. In the Ramachandran plots using MolProbity[48], no outliers were flagged.

Pull-down assays. GAI DELLA mutants were produced by site-directed mutagenesis[49] from pET47GAI6 (6× His fusion) to generate pET47GM1 (L31A), pET47GM2 (Y36A), pET47GM3 (V38A), pET47GM4 (M43A), pET47GM5 (E51R), pET47GM6 (E54R), pET47GM7 (L31A, V38A), pET47GM8 (Y36A, M43A) and pET47GM9 (E51R, E54R). Protein expression and preparation of cell extracts were performed as described previously without GA₃. Preparation of glutathione sepharose 4 beads (GE Healthcare) binding with glutathione *S*-transferase (GST)–GID1A and other pull-down procedures were followed as described previously[13] except for using anti-6× His antibodies (Santa Cruz) for detection of the GAI protein.

Trypsin digestion assay. GST–GID1A protein was purified by affinity chromatography (glutathione sepharose 4) and gel-filtration chromatography (Superdex 200 pg, GE Healthcare). GST–GID1A (80 μg ml⁻¹) and trypsin (2.5 μg ml⁻¹) were mixed in lysis buffer in the presence or absence of 100 μM GA₃ and the GAI DELLA domain at the room temperature (25 °C). The reaction was stopped by mixing with the SDS–PAGE sample buffer. The proteins were separated by 10% SDS–PAGE gel and stained by SimplyBlue SafeStain (Invitrogen). GST–GID1A was cleaved into GST and GID1A within a few minutes. MALDI-TOF MS of the resultant GID1A gave a peak at 37,204 Da (calculated 37,200 Da) corresponding to residues 14–345. GA-unbound GID1A is degraded rapidly within 10 min, whereas GA- and DELLA-bound GID1A were resistant for more than 20–30 min.

Circular dichroism measurements. The PCR-amplified GAI DELLA domain (11–97) was cloned into the BamHI/SacI site of pET-47b[+] to generate pET47G3. Protein expression and purification were performed as described previously. The protein sample solution contains 3 μg ml⁻¹ GAI DELLA in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM DTT. Circular dichroism spectra were recorded by a Jasco J-720W spectropolarimeter at 20, 37, 50 and 70 °C. Secondary structure estimations were calculated using the Jasco secondary-structure estimation software.

42. LeMaster, D. M. & Richards, F. M. 1H–15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–7268 (1985).

43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

44. Weeks, C. M. et al. Towards automated protein structure determination: *BnP*, the *SnB*-PHASES interface. *Z. Kristallogr.* **217**, 686–693 (2002).

45. Terwilliger, T. SOLVE and RESOLVE: Automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).

46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

47. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

48. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).

49. Weiner, M. P. et al. Site-directed mutagenesis of double-strand DNA by the polymerase chain reaction. *Gene* **151**, 119–123 (1994).

ARTICLES

HITS-CLIP yields genome-wide insights into brain alternative RNA processing

Donny D. Licatalosi[1], Aldo Mele[1], John J. Fak[1], Jernej Ule[3], Melis Kayikci[3], Sung Wook Chi[1], Tyson A. Clark[4], Anthony C. Schweitzer[4], John E. Blume[4], Xuning Wang[2], Jennifer C. Darnell[1] & Robert B. Darnell[1]

Protein–RNA interactions have critical roles in all aspects of gene expression. However, applying biochemical methods to understand such interactions in living tissues has been challenging. Here we develop a genome-wide means of mapping protein–RNA binding sites _in vivo_, by high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP). HITS-CLIP analysis of the neuron-specific splicing factor Nova revealed extremely reproducible RNA-binding maps in multiple mouse brains. These maps provide genome-wide _in vivo_ biochemical footprints confirming the previous prediction that the position of Nova binding determines the outcome of alternative splicing; moreover, they are sufficiently powerful to predict Nova action _de novo_. HITS-CLIP revealed a large number of Nova–RNA interactions in 3′ untranslated regions, leading to the discovery that Nova regulates alternative polyadenylation in the brain. HITS-CLIP, therefore, provides a robust, unbiased means to identify functional protein–RNA interactions _in vivo_.

The discovery of RNA molecules with catalytic activity[1] led to the hypothesis that, from the earliest life forms, RNA regulation evolved to have critical roles in living organisms[2–5]. Efforts to develop comprehensive understanding of protein–RNA interactions _in vivo_ have combined genetics, bioinformatics, microarray profiling and biochemical approaches. However, the latter have been hampered by methodological problems[6,7]; for example, co-immunoprecipitations can lead to re-association of protein–RNA complexes _in vitro_[8], non-specifically bound RNAs and additional co-precipitating RNA-binding proteins (RNABPs)[9].

We have taken a different approach towards understanding protein–RNA interactions by developing a crosslinking protocol that works in tissues, and can therefore be applied before protein purification. This method, termed CLIP[10,11], uses ultraviolet irradiation to induce covalent crosslinks between protein–RNA complexes _in situ_, allowing rigorous purification of RNABPs along with small fragments of RNA, which can be amplified and sequenced. CLIP has been used to study direct protein–RNA interactions extant in living cells[11–13], including identification of RNA targets[11] for the KH-type RNABP Nova[14,15], and the discovery of hnRNPA1-dependent regulation of a microRNA[12].

Genome-wide efforts to understand Nova function, using exon junction microarrays and bioinformatic analysis of Nova binding sites (YCAY clusters[15], characterized biochemically[16] and crystallographically[17]), suggested that the position of Nova binding to pre-mRNA predicted its action to enhance or inhibit alternate exon inclusion[18]. To identify direct Nova–RNA interactions _in vivo_, we applied high-throughput sequencing methods to CLIP. Here we demonstrate that this approach uncovers new biology in the brain, identifying functional interactions that mediate tissue-specific alternative RNA processing.

Genome-wide protein–RNA interaction maps

We studied Nova–RNA interactions in the mouse neocortex, which expresses the Nova2 protein[19]. We identified 2,481 Nova-bound RNAs (CLIP tags)[10,11] from five experiments using traditional CLIP strategies[10,11] and 412,686 CLIP tags from three experiments using

high-throughput pyrosequencing. Tags were filtered to eliminate those with imperfect (<80%) matches to genomic sequences, those with multiple genomic hits or those that were exact duplicates. The resulting set of 168,632 unique tags (≥92% bound to Nova2, and the remainder to either Nova1 or Nova2) included 123,734 tags mapping to messenger-RNA-encoding genes and 44,898 tags mapping to intergenic regions (Supplementary Fig. 2).

As negative controls, we repeated Nova CLIP using _Nova2_ knock-out brain or an irrelevant antibody, but were unable to amplify PCR products. We also sequenced 43,000 crosslinked RNA tags remaining after Nova immunoprecipitation, corresponding to a sample of all remaining RNABP–RNA interactions in the brain, and compared the frequency of Nova binding sites[15] in these control tags with that in Nova HITS-CLIP tags. Only the latter showed enrichment for YCAY sequences (observed:expected YCAY frequency was 3.56 for Nova tags, compared to 0.99 for the control tags, determined by Chi-squared distribution, $P < 10^{-227}$; see Methods), demonstrating the specificity of CLIP.

HITS-CLIP reproducibility and cluster analysis

We reasoned that one way to distinguish between biologically robust and transient Nova–RNA interactions would be to assess the reproducibility with which HITS-CLIP tags were identified in individual mice. Tags obtained from the neocortex of two P13 littermates showed a remarkable degree of similarity; when equal numbers were aligned across the entire mouse genome in 10-kilobase (kb) windows, a high correlation was evident both graphically (Fig. 1a) and statistically ($R^2 = 0.75$). To assess reproducibility more accurately, we focused on sites containing overlapping tags ('clusters'). A total of 19,156 clusters had at least 2 tags, 508 had 20 or more tags (Supplementary Fig. 3a), and 608 RefSeq transcripts had ≥6 clusters (Supplementary Fig. 2). Inter-animal clusters—sites containing at least one tag from each littermate—were highly reproducible; over 90% (9,697 out of 10,740) of sites containing tags from one animal also had at least one littermate tag. Finally, inspection of individual chromosomes and genes revealed that Nova clusters were highly
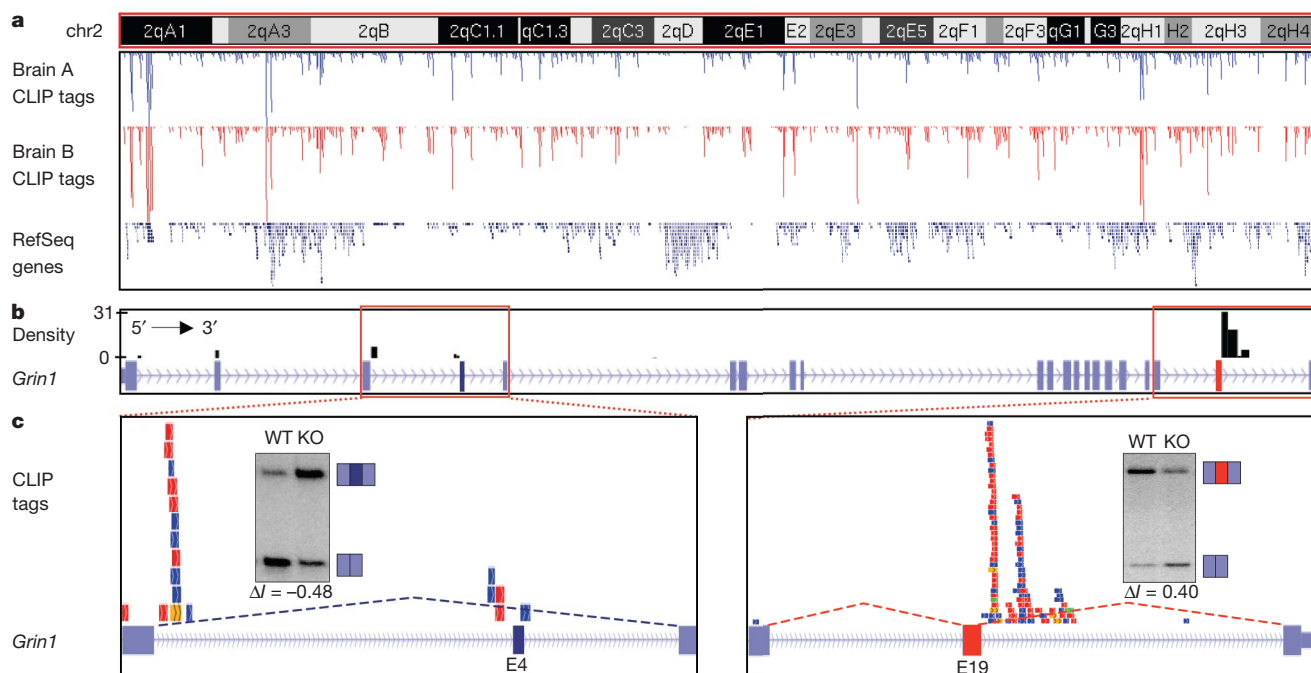
Figure 1 | HITS-CLIP genome-wide map of Nova–RNA binding sites.
a, Chromosome 2 (chr2) RefSeq genes and CLIP tags from the neocortex of two mouse littermates (brain A, 46,106 tags in 10,740 clusters; brain B, 100,874 tags in 15,805 clusters). **b**, Cluster density (number of tags per cluster length; black) in the *Grin1* transcript. **c**, *Grin1* E4 (left) and E19 (right) tags, one colour per biological replicate, predict Nova-dependent exon skipping and inclusion, respectively (experimental validation[20] is shown). KO, knockout; WT, wild type.

reproducible in both position and extent of crosslinking, and were specific to a subset of brain-expressed RefSeq genes (Fig. 1 and Supplementary Fig. 2).

To determine how faithfully Nova CLIP-tag clusters reflect previously defined consensus Nova binding sites[15], we analysed them for consensus motifs by MEME analysis, and found they were significant enriched in YCAY motifs (Supplementary Fig. 3). This was evident across all 19,156 Nova CLIP-tag clusters (3.9-fold; $P < 10^{-227}$), and in tags associated with functional Nova interactions (see below). Taken together, these observations indicate that HITS–CLIP reproducibly identifies discrete, YCAY-rich, Nova binding sites in mouse brain RNAs, and suggests that these binding sites may point to positions of functional Nova–RNA interactions.

HITS-CLIP analysis of splicing targets

HITS-CLIP offered a chance to compare predicted sites of Nova–RNA regulation derived from bioinformatic and microarray analysis[11,18,20] with interaction sites observed by *in vivo* crosslinking. A total of 39 previously validated[20] Nova2-regulated transcripts harboured Nova CLIP tags (ranging from 1 to 96 tags) within 3 kb of the alternative exon local region (bounded by the constitutive splice donor and acceptor exons) and 34 of these harboured CLIP-tag clusters. The position and YCAY content (4.1-fold enrichment; $P < 10^{-156}$) of these clusters was consistent with the predicted Nova bioinformatic map[18]. For example, YCAY-rich HITS-CLIP clusters were present downstream of the known Nova2 target *Grin1* exon 19 (E19; Fig. 1b, c (right))[20], in a position previously predicted by the Nova bioinformatic map[18] (Supplementary Fig. 4).

We also observed HITS-CLIP tags in *Grin1* upstream of an alternative exon (exon 4; E4) that was not a previously known Nova target. The position of these tags predicted Nova-dependent inhibition of E4 inclusion, which was confirmed experimentally (Fig. 1b, c (left)), suggesting that HITS-CLIP might provide a general means to identify new sites of protein–RNA regulation. Six additional transcripts with Nova HITS-CLIP clusters near regulated splice sites were tested; each was aberrantly spliced in *Nova2* knockout compared to wild-type

brain in a manner conforming to the Nova bioinformatic map (Supplementary Fig. 5).

To assess further how the position of Nova binding is related to the outcome of such splicing events, we analysed Nova HITS-CLIP tags in Nova-regulated exons newly identified using an updated version of exon-junction microarrays[20] harbouring probe sets for exon junctions in ~145,000 transcripts. Arrays were interrogated with RNA from wild-type or *Nova2* null neocortex, and results analysed with ASPIRE2, a revision of the ASPIRE algorithm[20] that searches for reciprocal changes in exon-included and exon-excluded probe sets. We identified 32 out of 45 previously validated[20] Nova2-dependent exons, as well as 46 new candidates with $|\Delta I|$ (the change in fraction of alternative exon inclusion[20]) values ranging from 0.19 to 0.60 and with characteristics seen previously[20] (Supplementary Fig. 6 and Supplementary Tables 1 and 2). To simplify subsequent analysis, we focused on 35 cassette exons, and confirmed that alternative splicing was Nova2-dependent in 7 out of 7 cassette exons assayed (Supplementary Fig. 4).

We generated a map in which we placed all 1,085 Nova CLIP tags identified from a total of 71 Nova2-regulated cassette exons (43 validated targets, and 28 newly predicted targets with $\Delta I > 0.2$ and ΔI-t-test > 25; see Methods) onto a single composite pre-mRNA (Fig. 2a and Supplementary Fig. 7). These tags spanned 11.5 kb, but were very heavily concentrated around splice sites, in positions that corresponded extremely well with the bioinformatically predicted Nova map[18], and with previous biochemical analysis of Nova-dependent splicing[21–23] (Fig. 2a). Furthermore, clusters in these regions showed a 3.4-fold enrichment in YCAY elements ($P < 10^{-174}$), with 72 of 123 clusters containing at least 3 YCAY elements within 30 nucleotides (nt), consistent with previous biochemical data[21–23].

We also noted some HITS-CLIP tags in unanticipated regions. For example, we observed frequent binding of Nova in intronic sequences upstream of Nova-regulated exons. However, binding to these sites was only robust in a limited number of transcripts (Fig. 2a and Supplementary Fig. 7). To generate a map representative of
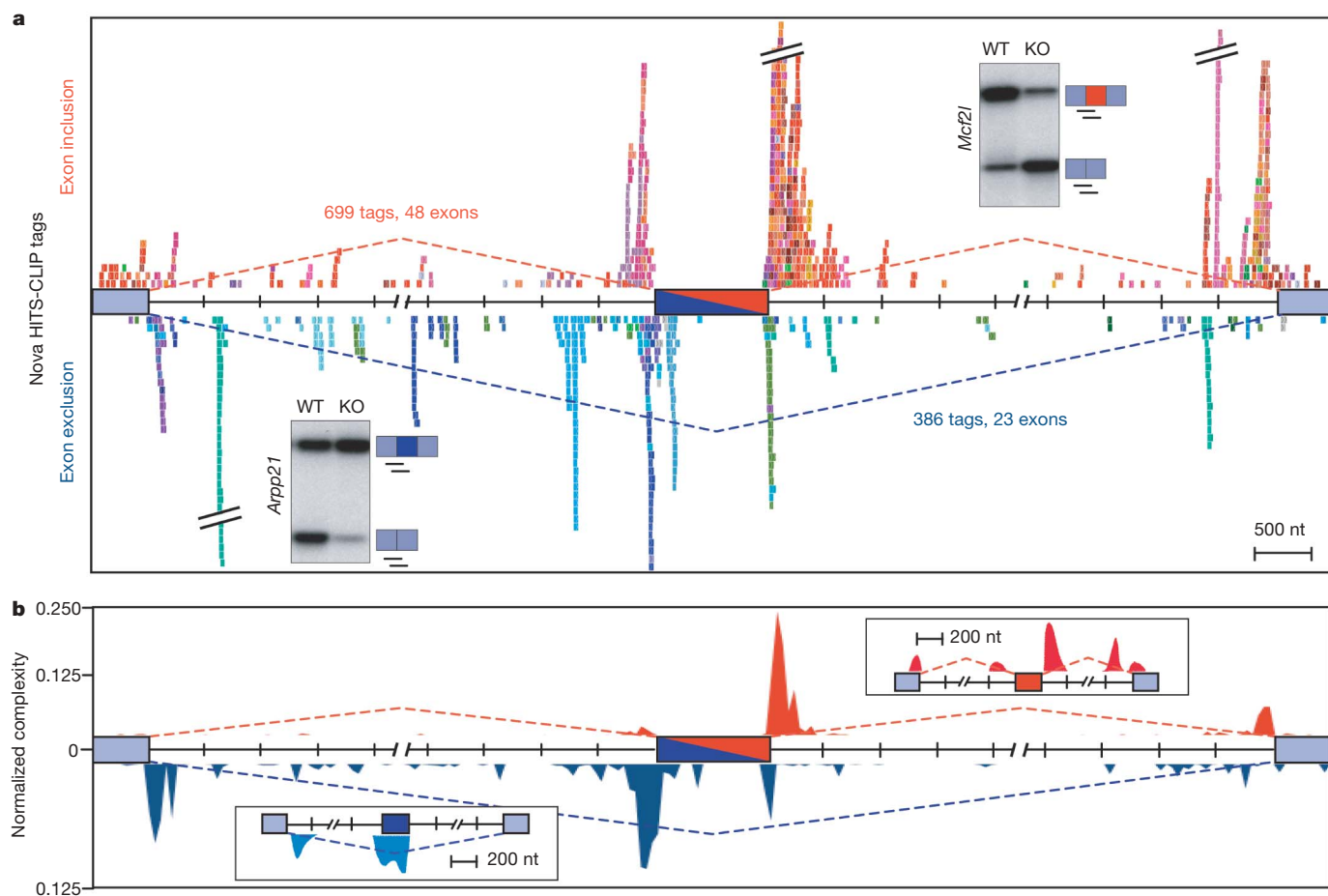
Figure 2 | Nova–RNA interaction maps associated with Nova-dependent splicing regulation. a, CLIP tags around all known Nova-regulated cassette exons, with one colour per transcript. Tags were mapped onto a composite transcript containing an alternative (dark blue/red box) and flanking constitutive (light blue box) exons. Tags are from transcripts showing Nova-dependent exon inclusion (top panel) or exclusion (bottom panel); representative examples of experimentally validated target RNAs (*Arpp21* and *Mcf2l*) are shown (insets). **b,** Normalized complexity map (see Methods) of Nova–RNA interactions recapitulate predicted maps[18] (insets) for Nova-dependent exon inclusion (red) or exclusion (blue).

consensus Nova action, we normalized our data, first to the number and distribution of CLIP tags between transcripts, and then to the number of different transcripts with tags at a given position (complexity). This allowed us to focus on potential regulatory binding sites common to several transcripts. This 'normalized complexity' map (Fig. 2b) demonstrated that Nova CLIP tags corresponded very precisely to the bioinformatically predicted sites of Nova action (Fig. 2b, insets). We conclude that HITS-CLIP confirms the hypothesis that Nova binding occurs directly on YCAY-rich elements near splice sites *in vivo*, and that the position of such Nova binding determines the outcome of Nova-dependent splicing regulation.

Nova regulates alternative polyadenylation

We next explored whether other HITS-CLIP clusters might reveal new Nova functions. Analysis of the genomic position of Nova clusters revealed that 23% of Nova HITS-CLIP tags mapped to intergenic regions (Fig. 3a). To examine the possibility that these tags may correspond to previously undescribed isoforms of RefSeq genes with alternative terminal exons, we examined the distance between intergenic clusters and neighbouring RefSeq genes. There was an exponential increase in the cumulative number of tags within 10 kb downstream of known stop codons, compared to a linear increase within 10 kb upstream of start codons (717 versus 101 clusters were present within 10 kb of the stop or start codon, respectively; Fig. 3b). This suggests that, in addition to binding known 3′ untranslated regions (UTRs; Fig. 3a), Nova binds to unannotated 3′ UTR extensions of known genes. Within 3′ UTRs, tags were enriched near

poly(A) sites, and to a lesser degree near stop codons (Fig. 3c). A large number of clusters were positioned within a few hundred nucleotides of poly(A) sites (Fig. 3d)—a region that contains core and potential auxiliary elements controlling transcript termination and poly(A) site use[24,25].

These observations suggested that Nova might function in a second pre-mRNA processing event in the mouse brain—regulated poly(A) site use (alternative polyadenylation)—a process about which little is known. We analysed alternative polyadenylation by hybridizing Affymetrix exon arrays with *Nova2* wild-type versus knockout brain RNA, and screened for changes in alternate 3′ UTR relative to total mRNA abundance (Supplementary Fig. 8). We identified 297 transcripts with such differences ($\geq$1.5-fold; $P < 0.05$); 43 contained 100 3′ UTR CLIP-tag clusters, and these were preferentially present near poly(A) sites (Fig. 3d).

We tested poly(A) site use in two candidates, *Cugbp2* and *Slc8a1*. Both have microarray-predicted Nova-dependent changes in 3′ UTR usage (1.5- and 2-fold, respectively), and both contained CLIP tags near poly(A) sites (Fig. 4a and Supplementary Fig. 9). RNase protection analysis (RPA) demonstrated that use of these poly(A) sites was increased in *Nova2* knockout brain (Fig. 4a, e and Supplementary Fig. 9); ΔC (the change in percentage of transcripts cleaved at the relevant poly(A) site, analogous to ΔI[18] for these transcripts was 0.22–0.25 (for example, 41% to 66% use of poly(A) site 2 (pA2) in *Cugbp2* transcripts in wild-type versus *Nova2* knockout brain; Fig. 4a), comparable in magnitude to Nova-dependent changes in alternative exon usage. Furthermore, the increase in proximal poly(A) use in *Cugbp2*
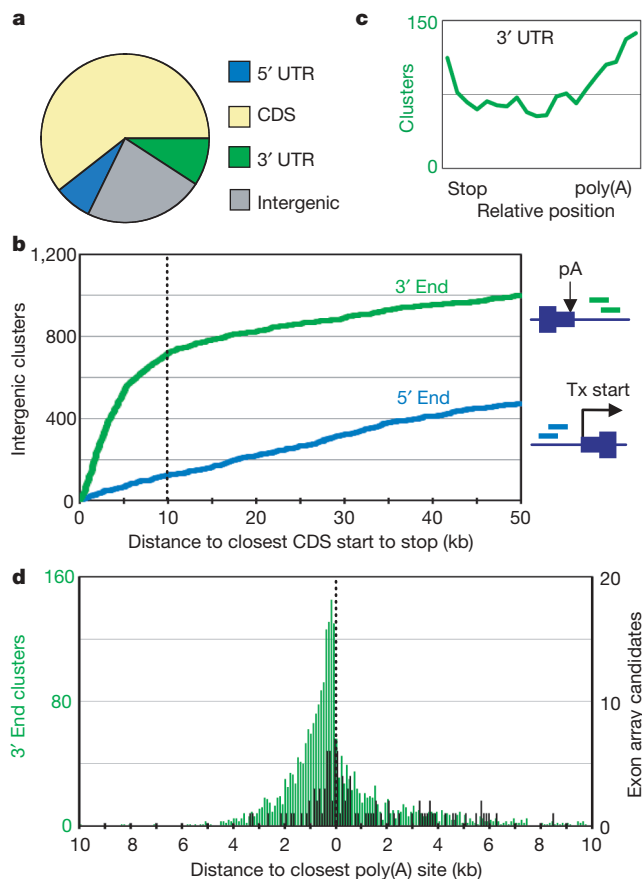
Figure 3 | Nova CLIP tags cluster near polyadenylation sites.
a, Distribution of Nova CLIP-tag clusters (RefSeq mm8; CDS refers to introns/exons between UTRs). **b**, Intergenic clusters were plotted relative to the closest annotated start (5′ end) or stop (3′ end) codon. pA, poly(A) site; Tx, transcription. **c**, Distribution of 2,465 clusters in 3′ UTRs (defined as distance from stop codon to closest transcript termination site). **d**, CLIP tag clusters relative to the closest poly(A) site (transcript end in UCSC known genes), plotted as clusters per 50 nt in 2,465 3′ UTRs (green bars) or in 43 Nova-dependent 3′ UTRs identified with exon arrays (black bars).

and *Slc8a1* transcripts in *Nova2* knockout brain was associated with reciprocal decreases in processing at distal poly(A) sites, suggesting that changes in the relative levels of alternatively polyadenylated *Cugbp2* and *Scl8a1* mRNAs are not due to differences in isoform stability, but result directly from aberrant poly(A) site use in the *Nova2* knockout.

We used quantitative polymerase chain reaction with reverse transcription (qRT–PCR) to measure the relative abundance of alternative poly(A) isoforms of 29 additional candidate Nova targets (from Fig. 3d). Twelve transcripts had significant changes in levels of alternatively polyadenylated transcripts ($P < 0.04$ in 11 out of 12; Fig. 4b). These transcripts did not change in overall transcript abundance (data not shown); nine were consistent with a Nova-dependent action to block, and three to enhance, use of the adjacent poly(A) site. Seventeen transcripts had either no change in poly(A) site usage in *Nova2* knockout brain (most of these were low abundance (at least in alternate 3′ isoforms)), and/or had confounding changes in overall steady-state transcript levels, and two transcripts had 3′ UTR changes as well as alternate splicing of terminal exons. We mapped Nova CLIP tags from the 12 Nova-regulated 3′ UTRs onto a composite transcript containing an alternative polyadenylation site (Fig. 4c), and onto a normalized complexity Nova–RNA 3′ UTR interaction map (Fig. 4d). This revealed reproducible Nova binding to discrete YCAY-rich (3.5-fold, $P < 10^{-227}$) regions flanking Nova-regulated alternative poly(A) sites. Taken together, the quantitative analysis of transcript levels and the HITS-CLIP map demonstrate that Nova
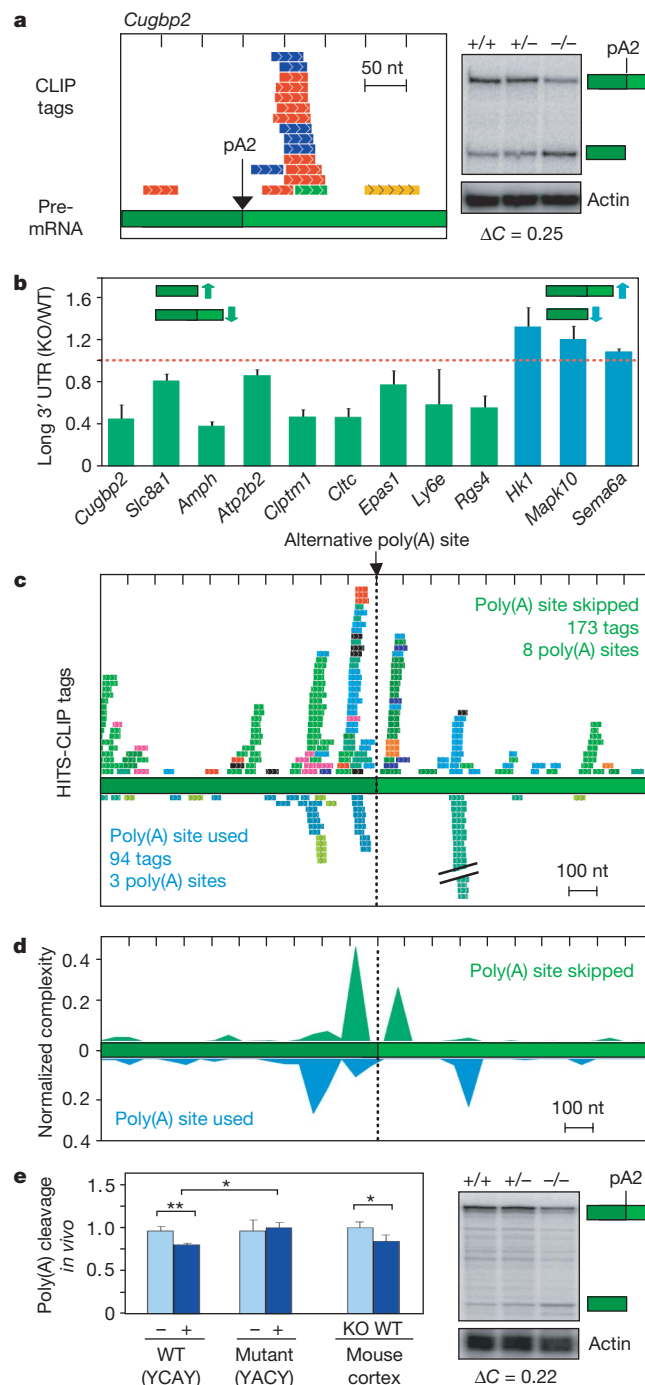
Figure 4 | Nova regulates alternative polyadenylation. a, Nova CLIP tags near *Cugbp2* pA2; each colour represents a biological replicate. Right: RPA measuring cleavage at pA2 in wild-type (WT, +/+), *Nova2* heterozygous (+/−) and knockout (KO, −/−) P10 neocortex, with actin as a normalization control and ΔC (see text) shown. **b**, qRT–PCR validation of Nova-dependent poly(A) regulation, using RNA from three WT or *Nova2* KO littermates, presented as long 3′ UTR isoform relative to total transcript abundance, normalized to WT (1.0; dotted line); error bars represent standard deviation. **c, d**, Composite (**c**) and normalized complexity (**d**) maps for Nova-regulated alternative polyadenylation sites (as in Fig. 2). **e**, Nova binds YCAY elements to directly regulate alternative polyadenylation. Left: qRT–PCR analysis of cells transfected with WT or mutant (three YACY mutations) poly(A) reporter constructs co-transfected with a control (−) or Nova2 expressing (+) plasmid, or WT versus *Nova2* KO neocortex (*$P < 0.02$, **$P < 0.01$); error bars represent standard deviation. Right panel: RPA for cleavage of *Slc8a1* pA2, otherwise as described in **a**.

binds to YCAY-rich elements flanking poly(A) sites and is necessary for their proper regulation in mouse brain.

To test whether Nova binding to these 3′ UTRs is sufficient to suppress poly(A) site use, we generated a green fluorescent protein (GFP) reporter construct containing parts of the *Slc8a1* 3′ UTR harbouring alternative poly(A) sites (Supplementary Fig. 9b), as well as mutant constructs in which YCAY elements were mutated to YACY (a sequence to which Nova does not bind[15]). Co-transfection of Nova2-expressing constructs with these reporters into 293T cells (which do not express Nova[22]) demonstrated a Nova- and YCAY-dependent reduction in alternative poly(A) site usage of the same magnitude and direction as seen in wild-type versus *Nova2* knockout neocortex (Fig. 4e). Taken together, these results identify direct Nova–RNA interactions in the 3′ UTR that regulate brain-specific alternative RNA processing.

Discussion

Genome-wide screens have been used to establish correlations between the action of RNABPs and biological diversity[6,7,26–30], but are unable to identify direct sites of RNA regulation. HITS-CLIP provides a general solution to this problem by generating a transcriptome-wide biochemical 'footprint' of protein–RNA interactions in living tissues. This in turn allows a direct comparison of predicted (for example, microarray or bioinformatically derived) and observed (HITS-CLIP) sites of action, and thereby provides a new platform for deriving functional RNABP maps and for assessing models of protein–RNA regulation.

HITS-CLIP extends our transcriptome-wide understanding of Nova–RNA interactions, which was previously limited to bioinformatic analysis of YCAY clusters within 200 nt of alternate or bounding constitutive exons[18]. Analysis of HITS-CLIP tags mapping to 71 Nova-regulated exons (Fig. 2) yielded a more refined map of Nova action. Over 91% of the normalized Nova binding associated with exon inclusion (Fig. 2b) occurred within 500 nt of either the alternative 5′ or the constitutive 3′ splice sites, whereas 74% of the normalized Nova binding associated with exon exclusion (Fig. 2b) occurred within 500 nt of the constitutive 5′ splice donor or surrounding the alternate exon. This strengthens the conclusion that the position of Nova–RNA interaction determines the outcome of splicing, an observation that may extend to splicing factors more generally[31]. Importantly, the strength of these correlations suggests that the HITS-CLIP map is sufficiently robust to predict protein–RNA regulation, as shown for seven new Nova splicing targets (Fig. 1d and Supplementary Fig. 5).

Whereas most of the Nova-regulated sites conform to a general set of rules based on direct Nova binding, there are also clear exceptions. For example, Nova binds robustly but in atypical positions in several regulated transcripts (for example, *Brsk2* and *Rap1gap*; Fig. 2a, b and Supplementary Fig. 7). Such examples may point to new mechanisms of Nova action, which may include interactions with other RNABPs. For example, Ptbp2 interacts with Nova and other RNABPs such as KSRP (also known as Khsrp) to modulate the outcome of alternative splicing[32,33]. In addition, RNA structure may regulate or be affected by interaction with other factors, as suggested by the ability of the splicing factor Mbnl1 to stabilize RNA hairpins[34], from analysis of splicing defects in *Mapt* that underlie frontotemporal dementia with Parkinsonism[35], and from structural studies of competition between hnRNP F and Ptbp1 binding to the *c-src* transcript (F. Allain, personal communication).

The unbiased nature of HITS-CLIP led to the unexpected identification of Nova binding near poly(A) sites and the recognition of its role in regulating alternative polyadenylation in the brain. The presence of such tissue-specific factors was postulated after the recognition of differential polyadenylation of immunoglobulin M heavy chain transcripts in B cells[36] and of *Calca* (calcitonin/calcitonin-related polypeptide) pre-mRNA in neurons[37]. Alternative poly(A) sites are present in ~50% of human genes[38], and their regulation is believed to have an important role in tissue and developmental mRNA regulation[39,40] as well as in human disease[41]. In particular,

brain mRNAs seem to be preferentially processed at promoter-distal poly(A) sites to generate long 3′ UTRs[42,43]. Interestingly, in 9 of 12 instances examined (Fig. 4b), Nova promoted the production of mRNAs with long 3′ UTRs. Thus, one important action of Nova may be to generate long 3′ UTRs in neurons, which may be subject to regulation by microRNAs or other RNABPs.

Numerous links have been made between pre-mRNA splicing and 3′-end processing[44,45], such as the observation that the splicing factor sex-lethal can regulate polyadenylation by competing with Cstf64 (also known as CstF2) for RNA binding[46]. Although we found two transcripts that had both Nova-dependent changes in splicing and polyadenylation, Nova can mediate splicing-independent alternative polyadenylation. For example, Nova suppresses the *Slc8a1* pA2 site in an intron-less transcript (Fig. 4), and *Cugbp2* and *Slc8a1* alternative polyadenylation was not coupled to alternative splicing in brain (unpublished data).

The Nova HITS-CLIP map offers insight into the mechanism of poly(A) site selection in the brain. Changes in the accessibility of core (for example, CPSF and CstF) or auxiliary factors to interact with *cis* elements surrounding the poly(A) site underlie the regulation of alternative polyadenylation[24,25,47,48]. We find no evidence that Nova regulates transcripts encoding such factors (including subunits of CPSF, CstF, CFI and CFII). Instead, our data point to Nova as a *trans*-acting factor that binds YCAY elements flanking regulated poly(A) sites, and that the position of Nova binding may determine whether it acts to promote or inhibit poly(A) site use (Fig. 4). For example, Nova CLIP tags overlap the canonical CPSF and/or CstF binding sites within 30 nt of the *Cugbp2* and *Slc8a1* poly(A) sites, which are suppressed by Nova. In contrast, in transcripts in which Nova enhances poly(A) site use, it binds to more distal elements, where it may antagonize the action of auxiliary factors. Therefore the position of Nova 3′ UTR binding may determine the outcome of poly(A) site selection in a manner analogous to its action on splicing regulation (Supplementary Fig. 1).

In summary, HITS-CLIP offers a powerful new platform for studying RNA regulation *in vivo*. This genome-wide biochemical approach complements bioinformatic, microarray and genetic studies. HITS-CLIP is able to identify biologically relevant interactions, providing a focus on direct protein–RNA contacts as critical points for understanding RNABP function. The unbiased nature of the platform holds the potential for new discovery, including the elucidation of preferred binding sequences and the identification of regulated RNA substrates. Identifying Nova as the first vertebrate factor to regulate alternative polyadenylation in mouse brain demonstrates that a single factor can regulate different aspects of tissue-specific RNA metabolism. Finally, the reproducible nature of HITS-CLIP suggests that it provides a robust platform to explore RNABP-dependent mechanisms of gene expression in complex and dynamic scenarios.

METHODS SUMMARY

HITS-CLIP. CLIP was performed on mouse *Nova2* wild-type and knockout (CD1) brains as described[11]. After PCR amplification, high-throughput sequencing was performed (454 Life Sciences).

Microarrays. For analysis of Nova-dependent alternate splice and alternate 3′ UTR variants, a custom exon junction array (Affymetrix) or MoEx 1.0 ST Affymetrix exon arrays, respectively, were used.

Bioinformatics. CLIP tags and clusters were analysed with BED or WIG formatted custom tracks using the UCSC Genome Browser and Genome Graph tools (http://genome.ucsc.edu). Composite maps were generated by determining the distance between tags and closest splice sites within the alternative exon local region and converted to coordinates in a BED format custom track, with tags from each gene assigned different colours. MEME sequence analysis was done using tools available at http://meme.sdsc.edu. ASPIRE2 was based on ASPIRE (ref. 20).

Biochemical and transfection assays. Biochemical assays were done using biological triplicate sibling mice, unless otherwise noted. RPAIII kits from Ambion were used, and RT–PCR experiments were done as described[21–23], with modifications described in Methods; qPCR was done with a MyIQ BioRad thermal cycler and data analysed as described in Methods. Wild-type or mutant GFP

alternative polyadenylation reporters were transfected into 293T cells in the presence or absence of pNova2 (described in Methods).

1. Zaug, A. J. & Cech, T. R. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231**, 470–475 (1986).
2. de Duve, C. Co-chairman's remarks: the RNA world: before and after. *Gene* **135**, 29–31 (1993).
3. Maizels, N. & Weiner, A. M. The 'last ribo-organism' was no breakthrough. *Nature* **330**, 616 (1987).
4. Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).
5. Sharp, P. A. On the origin of RNA splicing and introns. *Cell* **42**, 397–400 (1985).
6. David, C. J. & Manley, J. L. The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev.* **22**, 279–285 (2008).
7. Moore, M. J. & Silver, P. A. Global analysis of mRNA splicing. *RNA* **14**, 197–203 (2008).
8. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
9. Darnell, J. C., Mostovetsky, O. & Darnell, R. B. FMRP RNA targets: identification and validation. *Genes Brain Behav.* **4**, 341–349 (2005).
10. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
11. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
12. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
13. van der Brug, M. P. *et al.* RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways. *Proc. Natl Acad. Sci. USA* **105**, 10244–10249 (2008).
14. Buckanovich, R. J., Posner, J. B. & Darnell, R. B. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**, 657–672 (1993).
15. Darnell, R. B. Developing global insight into RNA regulation. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 321–327 (2006).
16. Buckanovich, R. J., Yang, Y. Y. & Darnell, R. B. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16**, 1114–1122 (1996).
17. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).
18. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
19. Yang, Y. Y. L., Yin, G. L. & Darnell, R. B. The neuronal RNA binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl Acad. Sci. USA* **95**, 13254–13259 (1998).
20. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nature Genet.* **37**, 844–852 (2005).
21. Jensen, K. B. *et al.* Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
22. Dredge, B. K. & Darnell, R. B. Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol. Cell. Biol.* **23**, 4687–4700 (2003).
23. Dredge, B. K., Stefani, G., Engelhard, C. C. & Darnell, R. B. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24**, 1608–1620 (2005).
24. Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485–1493 (2005).
25. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).
26. Ben-Dov, C., Hartmann, B., Lundgren, J. & Valcarcel, J. Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.* **283**, 1229–1233 (2008).
27. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
28. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* **8**, 749–761 (2007).
29. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
30. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
31. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
32. Markovtsov, V. *et al.* Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell. Biol.* **20**, 7463–7479 (2000).
33. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97**, 6350–6355 (2000).
34. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35**, 5474–5486 (2007).
35. Grover, A. *et al.* 5' splice site mutations in *tau* associated with the inherited dementia FTDP-17 affect a stem–loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **274**, 15134–15143 (1999).
36. Early, P. *et al.* Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* **20**, 313–319 (1980).
37. Rosenfeld, M. G. *et al.* Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. *Nature* **304**, 129–135 (1983).
38. Iseli, C. *et al.* Long-range heterogeneity at the 3' ends of human mRNAs. *Genome Res.* **12**, 1068–1074 (2002).
39. Edwalds-Gilbert, G., Veraldi, K. L. & Milcarek, C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* **25**, 2547–2561 (1997).
40. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
41. Danckwardt, S., Hentze, M. W. & Kulozik, A. E. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* **27**, 482–498 (2008).
42. Burge, C. B. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* doi:10.1038/nature07509 (this issue).
43. Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol.* **6**, R100 (2005).
44. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
45. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108**, 501–512 (2002).
46. Gawande, B., Robida, M. D., Rahn, A. & Singh, R. *Drosophila* Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J.* **25**, 1263–1272 (2006).
47. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**, 941–952 (1996).
48. Veraldi, K. L. *et al.* hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.* **21**, 1228–1238 (2001).
49. Clark, T. A. *et al.* Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* **8**, R64 (2007).

nature

METHODS

HITS-CLIP. After an initial RT–PCR using DNA primers complimentary to RNA linkers (previously described)[11], an additional PCR reaction was performed using the following fusion primers. AP5fusion1, 5′-GCCTCCCTCGCGCCATC-AGC**G**AGGGAGGACGATGCGG-3′. Five additional versions of this 5′ fusion primer were designed, each with a unique 'di-tag' (at the position underlined above; fusion primers 2–6 with AC, TA, CT, GC and GA, respectively), providing the ability to sequence multiple experiments simultaneously. In bold is the sequence complimentary to the 5′ RNA linker used. The remaining sequence at the 5′ end is that of the 454 Life Sciences 'Adaptor A.'

One 3′ fusion primer (BP3fusion) was designed consisting of a sequence complimentary to the 3′ RNA linker used (bold) and that of the 454 Life Sciences 'Adaptor B.' BP3fusion: 5′-GCCTTGCCAGCCCGCTCAG**CCGCT-GGAAGTGACTGACAC**-3′.

PCR amplification was performed using Accuprime Pfx (Invitrogen) and ranged between 10 and 15 cycles. The product was then run on a 2% agarose gel and purified using QiaEx II beads (Qiagen). A total of 100 ng of DNA was submitted for sequencing per run. The sequencing of CLIP tags was performed using 454 Life Sciences Adaptor A as the sequencing primer. The 454 adaptor sequences are as follows: Adaptor A, 5′-GCCTCCCTCGCGCCATC-3′; Adaptor B, 5′-GCCTTGCCAGCCCGCTCAG-3′.

Microarrays. RNA samples for all microarrays (Exon junction and MoEx 1.0 ST) were prepared using Whole Transcript Sense Target Labelling Assay and reagents (Affymetrix).

The research exon junction array contains probe sets for all exons and exon–exon junctions observed within transcripts in the input data. As input, the design uses transcript annotations from RefSeq (NCBI35), Ensembl (version 38) and ExonWalk. All transcripts were mapped to the August 2005 version of the mouse genome (NCBI 35, mm7). The array was designed primarily to interrogate well-annotated exons and splicing events from known genes. ExonWalk is a program that merges complementary DNA evidence together to predict full-length isoforms, including alternative transcripts. It was designed to incorporate the richness of transcript variation present in cDNA sequences but to limit some of the noise present in expressed-sequence-tag libraries by including several rules. ExonWalk requires that every exon and junction be: present in cDNA libraries of another organism, have multiple cDNA GenBank entries supporting it, or be evolving like a coding exon as determined by Exoniphy. More information on ExonWalk is available through the UCSC Genome Browser (http://genome.ucsc.edu/index.html).

Probe selection regions (PSRs) were created for the exon probe sets in a manner analogous to the Affymetrix exon arrays (http://www.affymetrix.com/support/technical/technotes/exon_array_design_technote.pdf). An exon was divided into multiple PSRs if there was evidence for alternative splice site usage. We targeted ten perfect match probes for each PSR. All PSRs greater than or equal to 25 bp were represented by at least one probe. All exon–exon junctions observed in the input transcripts (both alternative and constitutive junctions) are interrogated with a probe set containing eight perfect match probes that are tiled in one-base increments from the −4 to the +4 position (relative to the joining event).

Similar to the Affymetrix exon arrays (below), the probes on the array are all 25-base polymers and are designed for sense-strand target. The research junction array contains the same control and background probes as the Affymetrix exon arrays, such that data processing and array quality control methods can be shared by both array types.

Design statistics were as follows: Genome, mm7 (Aug 2005); transcript clusters (genes), 30,833; total transcripts, 145,993; observed junctions, 232,362; PSRs (exons), 270,632; and probes per junction, 8; probes per PSR (exon), 10.

Affymetrix MoEx 1.0 ST exon arrays for analysis of 3′ UTR expression were probed with neocortex RNA from four pairs of P10 wild-type and *Nova2* knockout littermate pairs. To identify candidates for Nova2-dependent alternative polyadenylation, core and extended probe sets mapping to, or within 10 kb of, RefSeq 3′ UTRs were first identified. A 3′ UTR index (analogous to the Splice Index)[49] was then calculated by normalizing the 3′ UTR probe set level to the core transcript level of the corresponding gene to identify differences in specific alternative isoforms, rather than differences due to changes in overall transcript level (see Supplementary Fig. 6 for additional information).

Bioinformatics. UCSC Genome tools (http://genome.ucsc.edu/index.html) were used extensively. To overlay CLIP tags from littermate animals in Supplementary Fig. 2c, because there were different numbers of CLIP tags in each experiment, we chose a computer-generated random subset of tags from the larger set to create sets with equal numbers of tags. A single outlier of intergenic tags (of unknown significance) on chromosome 12 was removed to generate more readable data, leaving a comparison of ~49,000 tags in each group. Graphs were generated with the UCSC Genome Browser or Genome Graph tools, after importing each tag set as a custom track.

YCAY analysis. To calculate YCAY enrichment, the number of observed and expected YCAYs (length of the sequence divided by 64) was calculated. For each data set, enrichment was determined by dividing the sum of observed YCAYs by the sum of expected YCAYs. To assess YCAY enrichment, a probability value was determined by calculating a Chi-squared distribution for each data set.

Normalized complexity map. To adjust for differences in CLIP tag numbers between different transcripts, the total tag number for each was normalized to 1.0. For each 50-nt window, the fraction of these tags present in individual transcripts was determined, and the sum of the fractions for all transcripts in each window was multiplied by the complexity (the number of different transcripts with a CLIP tag in the 50-nt window). As a result, regions with a high CLIP tag number and low complexity are minimized compared to regions with a smaller number of CLIP tags shared by multiple genes.

Biochemical and transfection assays. RPA, RT–PCR and qRT–PCR assays were performed using neocortex RNA from three pairs of P10 wild-type and *Nova2* knockout littermates. Complementary DNA of total RNA was generated using random hexamers and Superscript III (Invitrogen). For RT–PCR of each candidate splicing target, different cycle numbers were tested to ensure linear amplification of PCR products. Radiolabelled ^{32}P α-dCTP was added to the PCR reactions for the last two cycles. RPA probes were *in vitro* transcribed from linearized plasmids containing cloned sequences corresponding to the poly(A) sites in Cugbp2 (−275 to +68 nt) and Slc8a1 (−218 to +136 nt), and actin probes were synthesized from pTRI-Beta-Actin-mouse (Ambion). Radiolabelled probes were gel purified, and hybridizations and digestions performed as recommended by the manufacturer (Ambion RNAIII kit). RNase protected fragments and RT–PCR products were separated by electrophoresis on 6% polyacrylamide/7 M urea gels. qRT–PCR analysis of alternative polyadenylation candidates was performed using primers pairs located upstream and downstream of the poly(A) site being assayed. qRT–PCR analyses were performed in experimental triplicate using iTaq SYBR Green reaction mixes and iQ5 Real-Time PCR machines (BioRad).

The alternative poly(A) reporter plasmid was constructed by insertion of a PCR-amplified fragment corresponding to the Slc8a1 pA2 (−250 to +250 relative to the cleavage site) into plasmid vector pGFP (Clontech). The Slc8a1 pA3 (−1250 to +250) was inserted downstream of pA2. To ensure efficient expression in mammalian cells, a cytomegalovirus promoter was inserted upstream of the GFP open reading frame. YCAY elements were converted to YACY by PCR site-specific PCR mutagenesis. Wild-type or mutant poly(A) reporter plasmids were transfected (Lipofectamine-2000), with or without a Nova2-expressing plasmid (pNova2, R.B.D. *et al.*, unpublished), into 293T cells. All transfections were performed in triplicate. Primer sequences are available on request. For ASPIRE2, see Supplementary Information.

ARTICLES

Alternative isoform regulation in human tissue transcriptomes

Eric T. Wang[1,2]*, Rickard Sandberg[1,3]*, Shujun Luo[4], Irina Khrebtukova[4], Lu Zhang[4], Christine Mayr[5], Stephen F. Kingsmore[6], Gary P. Schroth[4] & Christopher B. Burge[1]

Through alternative processing of pre-messenger RNAs, individual mammalian genes often produce multiple mRNA and protein isoforms that may have related, distinct or even opposing functions. Here we report an in-depth analysis of 15 diverse human tissue and cell line transcriptomes on the basis of deep sequencing of complementary DNA fragments, yielding a digital inventory of gene and mRNA isoform expression. Analyses in which sequence reads are mapped to exon–exon junctions indicated that 92–94% of human genes undergo alternative splicing, ~86% with a minor isoform frequency of 15% or more. Differences in isoform-specific read densities indicated that most alternative splicing and alternative cleavage and polyadenylation events vary between tissues, whereas variation between individuals was approximately twofold to threefold less common. Extreme or 'switch-like' regulation of splicing between tissues was associated with increased sequence conservation in regulatory regions and with generation of full-length open reading frames. Patterns of alternative splicing and alternative cleavage and polyadenylation were strongly correlated across tissues, suggesting coordinated regulation of these processes, and sequence conservation of a subset of known regulatory motifs in both alternative introns and 3′ untranslated regions suggested common involvement of specific factors in tissue-level regulation of both splicing and polyadenylation.

The mRNA and protein isoforms produced by alternative processing of primary RNA transcripts may differ in structure, function, localization or other properties[1,2]. Alternative splicing in particular is known to affect more than half of all human genes, and has been proposed as a primary driver of the evolution of phenotypic complexity in mammals[3,4]. However, assessment of the extent of differences in mRNA isoform expression between tissues has presented substantial technical challenges[5]. Studies using expressed sequence tags have yielded relatively low estimates of tissue specificity, but have limited statistical power to detect differences in isoform levels[6–8]. Microarray analyses have achieved more consistent coverage of tissues[9], but are constrained in their ability to distinguish closely related mRNA isoforms. High-throughput sequencing technologies have the potential to circumvent these limitations by generating high average coverage of mRNAs across tissues while using direct sequencing rather than hybridization to distinguish and quantify mRNA isoforms[10,11].

Tissue-specific alternative splicing is usually regulated by a combination of tissue-specific and ubiquitously expressed RNA-binding factors that interact with *cis*-acting RNA elements to influence spliceosome assembly at nearby splice sites[1,2]. Many factors can both activate and repress splicing in different contexts, with activity often summarizable by an 'RNA map' describing dependence on the location of binding relative to that of core spliceosomal components[12,13].

A digital inventory of mRNA isoforms

To assess gene and alternative mRNA isoform expression, the mRNA-Seq protocol (Supplementary Methods) was used to amplify and sequence between 12 million and 29 million 32-base-pair (bp) cDNA fragments from ten diverse human tissues and five mammary epithelial

or breast cancer cell lines, generating over 400 million reads in total (Supplementary Fig. 1a). Tissue samples were derived from single anonymous unrelated individuals of both sexes; for one tissue, cerebellar cortex, samples from six unrelated men were analysed to assess variation between individuals (Supplementary Table 1). In total, ~60% of reads mapped uniquely to the genome, allowing up to 2 mismatches, and an additional 4% mapped uniquely to splice junctions. Thus, about two-thirds of reads could be assigned unambiguously to individual genes; the frequency of mapping to incorrect genomic locations was estimated to be ~0.1% (Supplementary Table 2).

Read density (coverage) was over 100-fold higher in exons than in introns or intergenic regions (Supplementary Fig. 1c), and only ~3% of reads mapped to ribosomal RNA genes, indicating that most reads derived from mature mRNA. Comparison of relative mRNA-Seq read densities to published quantitative polymerase chain reaction with reverse transcription (RT–PCR) measurements for 787 genes in two reference RNA samples[14] yielded a nearly linear relationship across ~5 orders of magnitude (Supplementary Fig. 1d), indicating that mRNA-Seq read counts give accurate relative gene expression measurements across a very broad dynamic range[10].

Alternative splicing is nearly universal

The mRNA-Seq data were used to assess the expression of alternative transcript isoforms in human genes, as illustrated for the mitochondrial phosphate transporter gene *SLC25A3* in Fig. 1a. Exons 3A and 3B of this gene are 'mutually exclusive exons' (MXEs), meaning that transcripts from this gene contain one or the other of these exons, but not both. Much greater read coverage of exon 3A was seen in heart and skeletal muscle, with almost exclusive coverage of exon 3B in

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden. [4]Illumina Inc., 25861 Industrial Boulevard, Hayward, California 94545, USA. [5]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [6]National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, New Mexico 87505, USA.
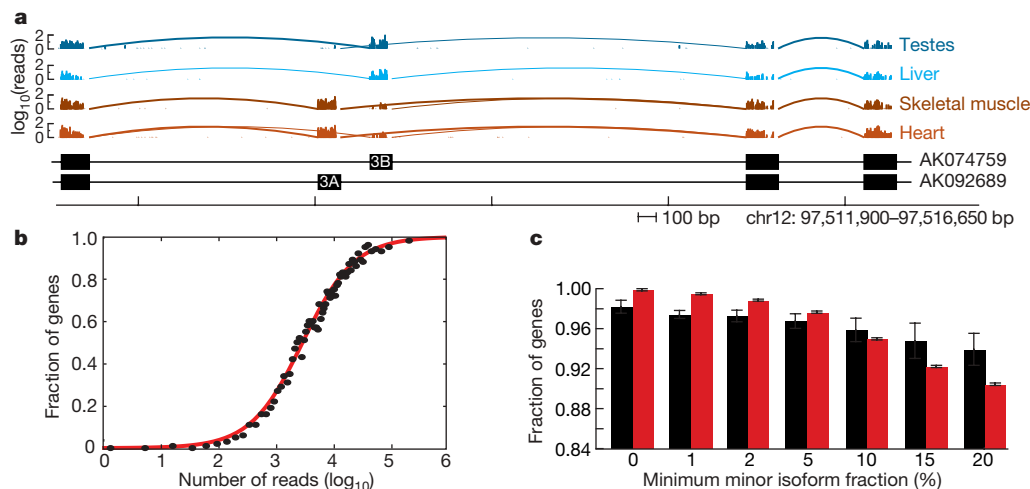*These authors contributed equally to this work.

Figure 1 | Frequency and relative abundance of alternative splicing isoforms in human genes. a, mRNA-Seq reads mapping to a portion of the *SLC25A3* gene locus. The number of mapped reads starting at each nucleotide position is displayed ($\log_{10}$) for the tissues listed at the right. Arcs represent junctions detected by splice junction reads. Bottom: exon/intron structures of representative transcripts containing mutually exclusive exons 3A and 3B (GenBank accession numbers shown at the right). **b,** Mean fraction of multi-exon genes with detected alternative splicing in bins of 500 genes, grouped by total read count per gene. A gene was considered as

alternatively spliced if splice junction reads joining the same 5′ splice site (5′SS) to different 3′ splice sites (3′SS) (with at least two independently mapping reads supporting each junction), or joining the same 3′SS to different 5′SS, were observed. The true extent of alternative splicing was estimated from the upper asymptote of the best-fit sigmoid curve (red curve). Circles show the fraction of alternatively spliced genes. **c,** Frequency of alternative splicing in the top bin (black bars) and after estimation (as in **b,** red bars), considering only events with relative expression of less abundant (minor) splice variant exceeding a given threshold. Error bars, s.e.m.

testes and liver (as well as in other tissues studied), consistent with the predominant heart and muscle symptoms of exon 3A mutation[15].

The genome-wide extent of alternative splicing was assessed by searching against known and putative splicing junctions using stringent criteria that required each alternative isoform to be supported by multiple independent splice junction reads with different alignment start positions. Binning the multi-exon genes in the RefSeq database (94% of all RefSeq genes) by read coverage and fitting to a sigmoid curve enabled estimation of the asymptotic fraction of alternatively spliced genes in this set as ~98% when excluding cell line data (Supplementary Fig. 2) and ~100% when using all samples (Fig. 1b). This analysis indicated that alternative splicing is essentially universal in human multi-exon genes, which comprise 94% of genes overall, with the important qualification that a portion of detected alternative splicing events may represent allele-specific splicing[16,17].

Some of these events may involve exclusively low frequency alternatively spliced isoforms. However, 92% of multi-exon genes were estimated to undergo alternative splicing when considering only events for which the relative frequency of the minor (less abundant) isoform exceeded 15% in one or more samples (Fig. 1c). Thus, 0.92 × 0.94 or ~86% of human genes were estimated to produce appreciable levels of two or more distinct populations of mRNA isoforms. Conversely, no evidence of alternative splicing was detected in the 6% of RefSeq genes annotated as consisting of a single exon, even when searching against junctions between predicted exons in these genes.

New exons and splice junctions not previously seen in transcript databases were identified by mapping the reads against predicted exons and junctions. This approach yielded a set of 1,413 high-confidence new exons (Supplementary Table 3), with an estimated false discovery rate (FDR) of <1.5% (Supplementary Information), and thousands of putative new splice junctions (not shown). Thus, mRNA-Seq has strong potential for discovery of new exons, although very substantial read depth is required to efficiently detect low-abundance isoforms (Supplementary Fig. 3).

Tissue-specific isoform expression

To explore the extent of tissue regulation of alternative transcripts, we examined eight common types of 'alternative transcript events'[1,2], each capable of producing multiple mRNA isoforms from human

genes through alternative splicing, alternative cleavage and polyadenylation (APA) and/or alternative promoter usage (Fig. 2). Event types considered included skipped exons and retained introns, in which a single exon or intron is alternatively included or spliced out of the mature message, and MXEs, described previously. Also included were alternative 5′ splice site (A5SS) and alternative 3′ splice site (A3SS) events, which are particularly difficult to interrogate by microarray analysis because the variably included region is often quite small. Tandem 3′ untranslated regions (UTRs) and alternative last exons (ALEs), in which alternative use of a pair of polyadenylation sites results in shorter or longer 3′ UTR isoforms or in distinct terminal exons, respectively, were also considered. Finally, we considered alternative first exons (AFEs), in which alternative promoter use results in mRNA isoforms with distinct 5′ UTRs.

For each of these event types, reads deriving from specific regions can support the expression of one alternative isoform or the other (Fig. 2). The 'inclusion ratio', defined as the ratio of the number of 'inclusion' (blue) reads to inclusion plus 'exclusion' (red) reads, can be used to detect changes in the proportions of the corresponding mRNA isoforms. The fraction of mRNAs that contain an exon—the 'per cent spliced in' (PSI or Ψ) value—can be estimated as the ratio of the density of inclusion reads (that is, reads per position in regions supporting the inclusion isoform) to the sum of the densities of inclusion and exclusion reads.

To assess tissue-regulated alternative splicing, a comprehensive set of ~105,000 events of these eight types was derived on the basis of available human cDNA and expressed sequence tag data. Reads supporting both alternative isoforms were observed for more than one-third of these events (Fig. 2), and the extent of tissue-specific regulation of these events was assessed by comparison of the inclusion ratio in each tissue relative to the other tissues, requiring a minimum of a 10% absolute change in inclusion ratio (Supplementary Fig. 4). Naturally, transcripts or isoforms identified as being differentially expressed between tissues will reflect the combined effects of cell-type-specific differences in transcript levels, variation in the relative abundances of cell types between tissues, and variations between the individuals from whom the tissues derived.

Notably, a high frequency of tissue-specific regulation was observed for each of the eight event types, including over 60% of the analysed skipped exon, A5SS, A3SS and tandem 3′ UTR events
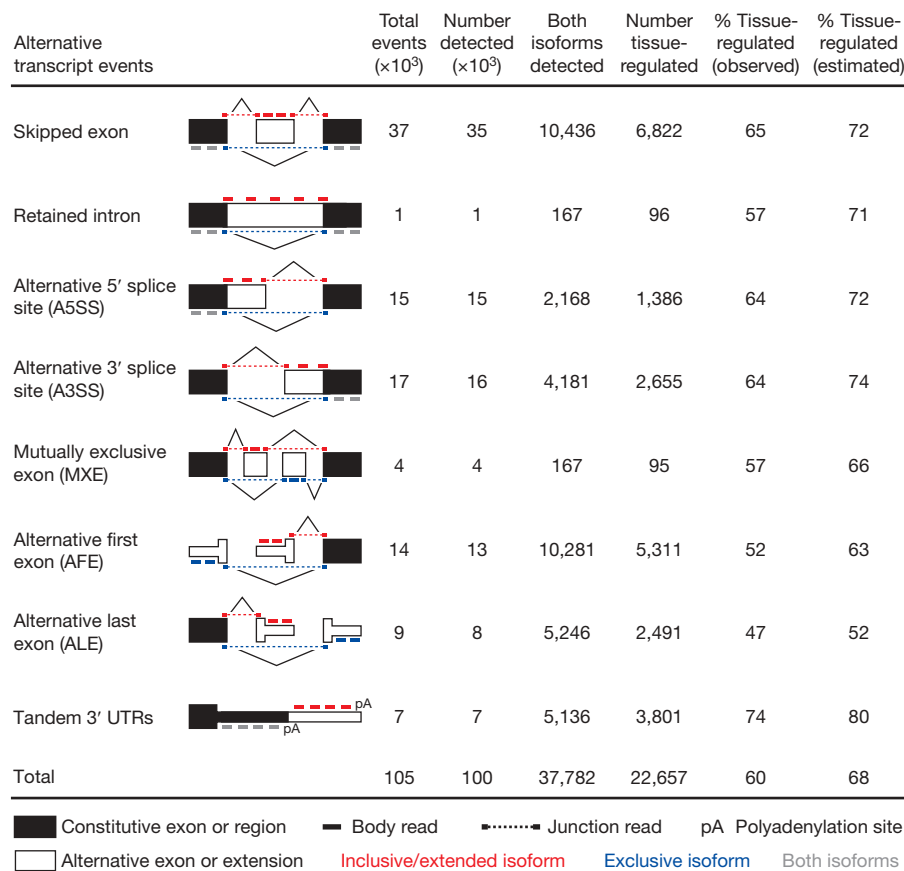
Alternative transcript events	Total events (×10³)	Number detected (×10³)	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon	37	35	10,436	6,822	65	72
Retained intron	1	1	167	96	57	71
Alternative 5′ splice site (A5SS)	15	15	2,168	1,386	64	72
Alternative 3′ splice site (A3SS)	17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE)	4	4	167	95	57	66
Alternative first exon (AFE)	14	13	10,281	5,311	52	63
Alternative last exon (ALE)	9	8	5,246	2,491	47	52
Tandem 3′ UTRs	7	7	5,136	3,801	74	80
Total	105	100	37,782	22,657	60	68

■ Constitutive exon or region   ▬ Body read   ┅┅ Junction read   pA Polyadenylation site
□ Alternative exon or extension   Inclusive/extended isoform   Exclusive isoform   Both isoforms

Figure 2 | Pervasive tissue-specific regulation of alternative mRNA isoforms. Rows represent the eight different alternative transcript event types diagrammed. Mapped reads supporting expression of upper isoform, lower isoform or both isoforms are shown in blue, red and grey, respectively. Columns 1–4 show the numbers of events of each type: (1) supported by cDNA and/or EST data; (2) with ≥1 isoform supported by mRNA-Seq reads; (3) with both isoforms supported by reads; and (4) events detected as tissue-regulated (Fisher's exact test) at an FDR of 5% (assuming negligible technical variation[10]). Columns 5 and 6 show: (5) the observed percentage of events with both isoforms detected that were observed to be tissue-regulated; and (6) the estimated true percentage of tissue-regulated isoforms after correction for power to detect tissue bias (Supplementary Fig. 6) and for the FDR. For some event types, 'common reads' (grey bars) were used in lieu of (for tandem 3′ UTR events) or in addition to 'exclusion' reads for detection of changes in isoform levels between tissues.

(Fig. 2 and Supplementary Table 4). In all, a set of over 22,000 tissue-specific alternative transcript events was identified, far exceeding previous sets of tissue-specific alternative splicing events that have typically numbered in the hundreds to low thousands[6–9,18,19]. Tissue-regulated skipped exon and MXE events are listed in Supplementary Tables 5 and 6, respectively. Binning events by expression level commonly yielded sigmoid curves for the fraction of tissue-regulated events of each type, enabling estimation of the true frequency of tissue regulation for each event type (Supplementary Figs 5 and 6). These estimates, ranging from 52% to 80% (Fig. 2), indicated that most alternative splicing events are regulated between tissues, providing an important element of support for the hypothesis that alternative splicing is a principal contributor to the evolution of phenotypic complexity in mammals.

Individual-specific isoform expression

To assess the extent of alternative splicing isoform variation between individuals in comparison to tissue-regulated alternative splicing, the correlations among the vectors of inclusion ratios for all expressed skipped exons between pairs of samples were determined (Fig. 3); this was performed similarly for other event types (not shown). In this analysis, strong clustering of the six cerebellar cortex samples was observed, with generally higher correlations among these samples than between pairs representing distinct tissues. Strong clustering of the five cell lines was also observed. This probably results from a combination of factors, including the common mammary epithelial origin of the cell lines studied, similar adaptations to culture conditions, and the high diversity of the tissues chosen.

The extent of variation in alternative isoform expression between individuals was also addressed by determining the number of differentially expressed exons among the six cerebellar cortex samples. Using the same approach as in Fig. 2, between ~10% and 30% of alternative transcript events showed individual-specific variation, depending on the event type (Supplementary Fig. 7), providing updated estimates of the scope of mRNA isoform variation between individuals[16]. These numbers are higher than estimates based on microarray analyses[20], but are in general agreement with an integrated analysis of multiple data types that estimated that ~21% of alternatively spliced genes are affected by polymorphisms that alter the relative abundances of alternative isoforms[17]. However, these frequencies are still below the 47–74% of events that showed variation among the ten tissues (Fig. 2), and approximately twofold to threefold less than the frequencies observed in comparisons among subsets of six tissues (Supplementary Fig. 7), indicating that, although inter-individual variation is fairly common, it is still substantially less frequent than variation between tissues. Thus, most of the differences observed between tissue samples are likely to represent tissue-specific rather than individual-specific variation.

Switch-like alternatively spliced exons

The quantitative nature of the mRNA-Seq approach allowed assessment of both subtle and switch-like alternative splicing events. By
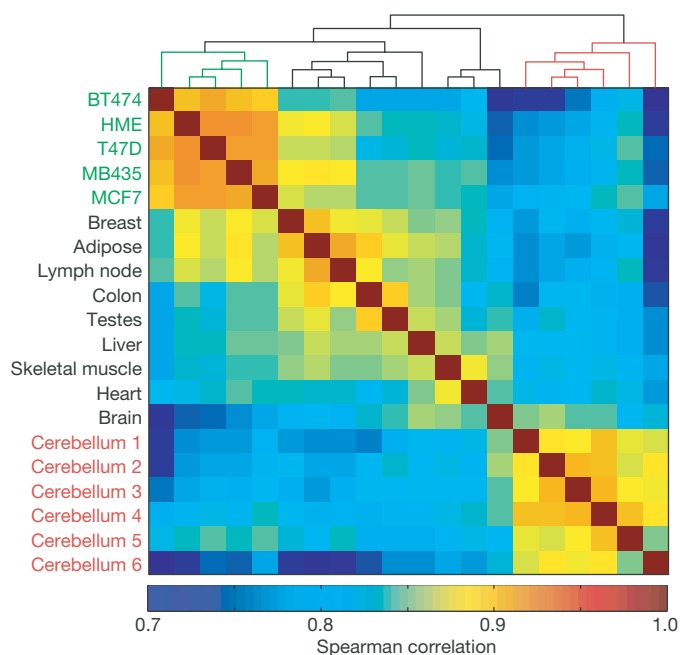
Figure 3 | The extent of individual-specific differences in alternative isoform expression. Spearman correlations of inclusion ratios for skipped exons in human tissues and cell lines (see Methods Summary). Correlations were computed separately for each pair of tissues and cell lines, and clustered according to similarity using average linkage hierarchical clustering.

comparing inclusion levels of skipped exons between tissues, a class of 'switch-like' exons was observed that had markedly different inclusion levels between different tissues (shown for heart versus nine other tissues in Fig. 4a). The examples shown in colour in Fig. 4a (for

example, *TPM1* exon 2, with Ψ of 2% in heart and 95% in skeletal muscle, and the *SLC25A3* MXE pair shown in Fig. 1a) underscore the flexibility of the splicing regulatory machinery, with a sizeable number of exons being recognized predominantly as exons in one tissue and predominantly as introns in another tissue, even for developmentally related pairs of tissues such as heart and skeletal muscle.

To characterize functional features of such switch-like exons, skipped exons and MXEs were divided into groups depending on their 'switch score', defined as the maximum pairwise Ψ difference between tissues. Switch scores for pairs of MXEs were shifted towards higher values relative to skipped exons ($P = 3.7 \times 10^{-5}$, Kolmogorov–Smirnov test; Fig. 4b), suggesting that MXEs are more often involved in regulating highly tissue-specific functions. Preservation of the reading frame in both isoforms was observed more commonly for exons with higher switch scores both for skipped exons, consistent with ref. 19, and to an even greater extent for MXEs (Fig. 4c). Thus, switch-like regulation seems to be used preferentially to express distinct 'full-length' protein isoforms in different tissues rather than as a means to switch off genes through production of truncated proteins or of messages subject to nonsense-mediated mRNA decay[21]. Indeed, genes containing skipped exons with high switch scores were enriched for Gene Ontology functional categories including 'developmental processes', 'cell communication', 'signal transduction' and 'regulation of metabolism' that are likely to contribute to fundamental differences in the biology of different human tissues (Supplementary Table 7).

Notably, skipped exons with switch scores exceeding 0.5 showed higher sequence conservation in the regulated exon itself[19] and in portions of the flanking introns than exons with lower switch scores (Fig. 4d). This observation suggested that such exons are of unusual biological importance and that switch-like regulation between tissues requires the presence of additional splicing regulatory sequence information, particularly in adjacent intronic regions.
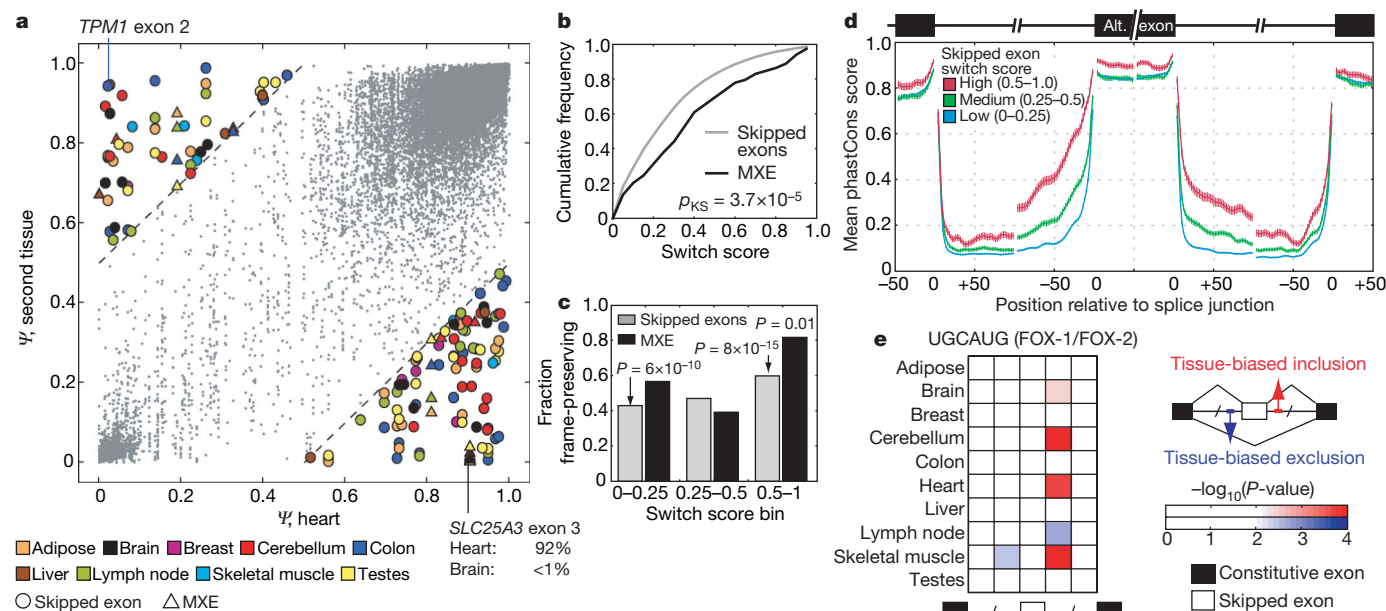
Figure 4 | Conservation and function of switch-like alternative splicing exons. a, Scatter plot showing Ψ values of skipped exons and MXEs for which switch score was determined on the basis of comparison of heart (*x*-axis) with a second tissue (*y*-axis). Exons with a switch score >0.5 are shown as filled symbols; others are shown as small grey dots. **b,** Cumulative distribution functions of switch scores for skipped exons and MXE pairs (*P*-value based on Kolmogorov–Smirnov test). **c,** Reading frame preservation of skipped exons and MXEs grouped by switch score. Skipped exons with lengths divisible by 3 and MXE pairs with lengths differing by 0 or a multiple

of 3 were considered to preserve the reading frame. *P*-values are based on Fisher's exact test. **d,** Conservation in skipped exon and flanking intron regions grouped by skipped exon switch score. The mean per-position phastCons score (from alignment of four mammalian genomes) and s.e.m. are shown. **e,** Enrichment of UGCAUG motifs near tissue-regulated skipped exons. Coloured squares represent $-\log_{10}(P\text{-value})$ for the enrichment of UGCAUG counts relative to cohorts of control hexanucleotides in regions surrounding skipped exons with significantly increased (red) or decreased (blue) inclusion in each tissue with respect to other tissues.

FOX-1 and FOX-2 activity map

Among the best-characterized tissue-specific splicing factors are the FOX-1 (also known as A2BP1) and FOX-2 (RBM9) proteins, which bind RNA *cis*-elements that contain UGCAUG hexanucleotides or closely related sequences[22–24]. Analysis of UGCAUG frequencies revealed substantial enrichment in the intron immediately downstream of exons with increased inclusion in heart, skeletal muscle, brain and cerebellar cortex (Fig. 4e)—tissues where FOX proteins are highly expressed, suggesting common splicing activation activity in this location[22–24]. Enrichment of UGCAUG hexanucleotides was also noted upstream of exons that had reduced inclusion in skeletal muscle, suggesting possible repressive activity in this context. This example illustrates the power of these expanded tissue-specific exon sets for inference of 'tissue RNA maps', summarizing both the location-dependent activity and tissue specificity of splicing regulatory elements.

Applying a similar approach to analyse enrichment of all hexanucleotides in regions adjacent to tissue-specific exons identified 362 motif/tissue enrichment patterns (at an estimated 17% FDR), representing hexanucleotides that showed significant enrichment adjacent to exons with increased or decreased inclusion in specific cell lines or tissues (Supplementary Table 8). Enrichment of UGCAUG downstream of exons with high inclusion in skeletal muscle appears as the third most significant motif/tissue pair, after enrichment of UCUCUC and CUCUCU (resembling the binding motifs of PTBP1 (also known as PTB) and PTBP2 (nPTB)[25]) upstream of exons with increased inclusion in cerebellar cortex. The remaining motif/tissue pairs contained a variety of known regulatory elements, including ACUAAC (see later), as well as putative new regulatory motifs.

Coordination of splicing and polyadenylation

Tandem 3′ UTR events showed an even higher frequency of tissue-regulated expression than skipped exons or other alternative splicing events studied (Fig. 2), yet little is known about how tissue regulation of tandem UTRs is accomplished (for example, whether through APA or through the differential stability of alternative UTR isoforms). By grouping tandem 3′ UTRs by switch score, the most switch-like events showed increased sequence conservation relative to events with lower switch scores in the vicinity of and upstream of the proximal (5′) polyadenylation signal (PAS), and also upstream of the distal (3′) PAS (Fig. 5a). Whereas *cis*-regulatory elements contributing to differential stability should be located predominantly in the region unique to the long UTR isoform, APA could be regulated by elements located near to either or both PASs. The observation of increased conservation around and upstream of the proximal PAS in switch-like tandem UTRs therefore supports a primary role for regulation at the level of APA.

In assessing the spectrum of *cis*-elements that may drive tissue regulation of tandem 3′ UTRs, a set of heptanucleotides was identified that showed high conservation in the extension region of tandem 3′ UTRs (Fig. 5a, inset), with signal:background ratios in four mammals[26] exceeding 2:1. As expected, this set included the extended (seven-base) seed matches to a number of conserved mammalian microRNAs (miRNAs)[26–28]. Surprisingly, it also included all eight of the heptanucleotides that contain the FOX-1/FOX-2 consensus binding motif, UGCAUG: all such heptanucleotides had signal/background ratios above 2.5:1, exceeding the signal/background ratio observed for seed matches to important miRNAs such as miR-7 and miR-181 (inset, Supplementary Table 9). Strong conservation of UGCAUG motifs in this location (>1 kilobase on average from the nearest splice site) would not be expected on the basis of the canonical splicing regulatory activity of FOX-1/FOX-2 proteins. Instead, the high conservation observed in extended 3′ UTR regions suggests that these factors (or others with identical RNA-binding specificity) have additional 3′ UTR-related roles, for example, in APA or in mRNA localization and/or translation.

To investigate possible connections between tissue-specific regulation of alternative splicing and APA further, global patterns of

tissue-specific alternative isoform expression were compared. By applying singular value decomposition (SVD) (Supplementary Methods) to the vectors of inclusion ratios across samples for each alternative splicing and APA event type separately, a strong and consistent separation of the breast cell lines (four cancer-derived and one immortalized cell line) from all tissue samples was observed (Fig. 5c,
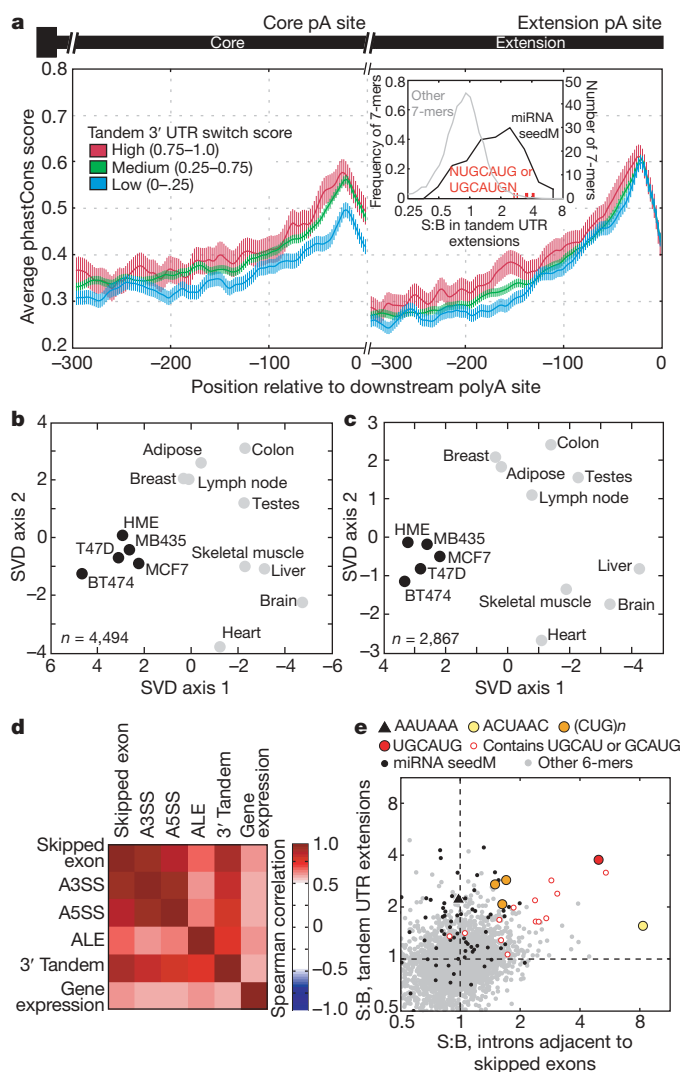
Figure 5 | Evidence for coordination between splicing and polyadenylation. a, Mean and s.e.m. of per-position phastCons score in the region 300 bp upstream of proximal and distal cleavage sites for tandem 3′ UTRs grouped by switch score. Inset, increased conservation of FOX-1/FOX-2 motifs in tandem 3′ UTR extension regions. All non-CpG-containing heptanucleotides (grey line), miRNA seed matches (black), and 7-mers containing UGCAUG (red) are shown. **b**, SVD analysis of skipped exon inclusion ratio values across tissues and cell lines for skipped exons meeting minimum read coverage criteria in each of the 14 samples. Projections are shown in the dimensions corresponding to the two leading eigenvalues, which accounted for 25% of the variance. **c**, SVD analysis of tandem UTR inclusion ratio values (as in **b**). **d**, SVD analysis was conducted for the 14 samples on the basis of inclusion ratio values for the five indicated alternative transcript event types or on the basis of gene expression values. Spearman correlations between corresponding pairwise distances in projections of the sort shown in **b** and **c** are shown. **e**, Signal:background (S:B) ratios of non-CpG-containing hexanucleotides in introns flanking skipped exons (*x*-axis) and in extended 3′ UTR regions (*y*-axis). The canonical PAS hexanucleotide AAUAAA (black triangle), hexanucleotides corresponding to seed matches to conserved mammalian miRNAs (black dots), hexanucleotides corresponding to binding motifs for the indicated splicing or 3′ UTR-binding factors (coloured), and other hexanucleotides (small gray dots) are shown.

d). This separation implied the existence of a systematic difference in RNA processing regulation between cell lines and tissues that held for all types of alternative events studied. For most alternative splicing and APA events, SVD analysis yielded similar groupings of tissues, for example, with heart, skeletal muscle, brain and liver consistently clustered (Supplementary Fig. 8). Consistent with this observation, pairwise distances between SVD projections for different types of alternative splicing events, for example, skipped exons, A5SS and A3SS events, were all highly correlated (Fig. 5e), suggesting similarities in the regulatory control of these types of events[1,2,13]. More surprisingly, distances between SVD projections for tandem 3′ UTR events also correlated highly with distances for events controlled purely at the level of splicing such as skipped exons (Fig. 5e). This observation raised the possibility that splicing and polyadenylation may be coordinately regulated across human tissues.

To explore possible regulatory connections between splicing and polyadenylation regulation (for example, refs 29–32), the conservation of hexanucleotides adjacent to conserved alternative splicing and APA events was compared. Whereas canonical 3′ UTR regulatory motifs such as the consensus PAS hexanucleotide AAUAAA and various miRNA seed matches showed high signal:background ratios, often 1.5:1 or higher, in extended 3′ UTR regions, these motifs generally had signal/background ratios close to 1:1 in alternatively spliced introns. However, a distinct subset of motifs with high signal:background ratios in both UTRs and introns was also observed, several of which corresponded to well-known splicing-related motifs (Fig. 5h and Supplementary Table 9). This set included not only the FOX-1/FOX-2 motif UGCAUG and variations, consistent with the heptanucleotide analysis of Fig. 5a, but also permutations of $(CUG)_n$, which represent putative substrates of the bruno-like (BRUNOL, also known as CELF) and muscleblind-like (MBNL) families of muscle- and brain-specific splicing factors[33]. The highly significant signal:background ratio in both 3′ UTRs and introns suggested that these well-known splicing-related motifs also commonly have 3′ UTR-related roles—for example, control of APA or of mRNA stability, localization or translation—as recently demonstrated for the NOVA family of splicing factors[34].

The hexanucleotide ACUAAC, an excellent match to the consensus binding motifs of STAR family RNA-binding factors, in particular quaking homologue (QKI)[35], was also notable. Not only did ACUAAC have significant signal:background ratio in 3′ UTRs, as expected from the known role of QKI in control of mRNA stability[36], but it also showed an extremely high signal:background ratio in introns, exceeding 7:1. This extreme conservation suggested a common and important function in splicing regulation—a role that has been suggested but not yet directly demonstrated[9,37]. Motif enrichment analyses also suggested a possible role in brain-specific APA regulation (Supplementary Fig. 9).

Discussion

We conclude that the coordination between tissue-specific alternative splicing and APA events implied by the correlated patterns of tissue bias observed in Fig. 5 may be mediated at least in part by tissue-specific RNA-binding factors that have roles in regulation of both of these RNA processing steps. Such factors may include both canonical tissue-specific splicing factors (for example, of the FOX-1/FOX-2 and CELF families), moonlighting in 3′ UTR-related roles, and also canonical UTR-binding factors such as QKI. Such functional duality has the potential to enable tightly coordinated regulation of polyadenylation and splicing, ensuring that the appropriate UTR regulatory sequences are expressed in conjunction with the coding regions for the relevant tissue-specific protein isoforms.

METHODS SUMMARY

Tissues and cell lines. Tissue samples from individual unrelated anonymous donors (Supplementary Table 1) were obtained from Ambion for the following tissue types: adipose, whole brain, breast, colon, heart, liver, lymph node, skeletal muscle and testes. Cerebellar cortex samples were obtained from six anonymous unrelated donors, according to NIH guidelines for confidentiality and privacy using protocols described previously[38]. HME is a human mammary epithelial cell line immortalized with human TERT[39]. The other cell lines are all breast cancer cell lines derived from invasive ductal carcinomas (ATCC). MCF-7, BT474 and T47D are oestrogen-receptor- and progesterone-receptor-positive; MDA-MD435 is negative for both.

Library preparation for Illumina sequencing. Poly-T capture beads were used to isolate mRNA from 10 µg of total RNA. First-strand cDNA was generated using random hexamer-primed reverse transcription, and subsequently used to generate second-strand cDNA using RNase H and DNA polymerase. Sequencing adaptors were ligated using the Illumina Genomic DNA sample prep kit. Fragments ~200 bp long were isolated by gel electrophoresis, amplified by 16 cycles of PCR, and sequenced on the Illumina Genome Analyser, as described[40,41].

Computational analyses of mRNA-Seq read data. Computational and statistical methods used in analysis of the read data are described in the Supplementary Methods. High-confidence new exons were required to be supported by at least one splice juction read involving each splice site, and at least one exon body read; putative new splice junctions required splice junction read support only[42].

1. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
2. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
4. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
5. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
6. Xu, Q., Modrek, B. & Lee, C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754–3766 (2002).
7. Gupta, S., Zink, D., Korn, B., Vingron, M. & Haas, S. A. Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. *BMC Genomics* **5**, 72 (2004).
8. Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biol.* **5**, R74 (2004).
9. Sugnet, C. W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLOS Comput. Biol.* **2**, e4 (2006).
10. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
11. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
12. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
13. Wang, Z., Xiao, X., Van Nostrand, E. & Burge, C. B. General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* **23**, 61–70 (2006).
14. Shi, L. *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnol.* **24**, 1151–1161 (2006).
15. Mayr, J. A. *et al.* Mitochondrial phosphate-carrier deficiency: a novel disorder of oxidative phosphorylation. *Am. J. Hum. Genet.* **80**, 478–484 (2007).
16. Graveley, B. R. The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population. *Trends Genet.* **24**, 5–7 (2008).
17. Nembaware, V., Wolfe, K. H., Bettoni, F., Kelso, J. & Seoighe, C. Allele-specific transcript isoforms in human. *FEBS Lett.* **577**, 233–238 (2004).
18. Pan, Q. *et al.* Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16**, 929–941 (2004).
19. Xing, Y. & Lee, C. J. Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS Genet.* **1**, e34 (2005).
20. Kwan, T. *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nature Genet.* **40**, 225–231 (2008).
21. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
22. Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25**, 10005–10016 (2005).
23. Auweter, S. D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
24. Nakahata, S. & Kawamoto, S. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res.* **33**, 2078–2089 (2005).
25. Oberstrass, F. C. *et al.* Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–2057 (2005).

26. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).

27. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).

28. Majoros, W. H. & Ohler, U. Spatial preferences of microRNA targets in 3′ untranslated regions. *BMC Genomics* **8,** 152 (2007).

29. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416,** 499–506 (2002).

30. McCracken, S., Lambermon, M. & Blencowe, B. J. SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.* **22,** 148–160 (2002).

31. Castelo-Branco, P. *et al.* Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* **24,** 4174–4183 (2004).

32. Zhang, L., Lee, J. E., Wilusz, J. & Wilusz, C. J. The RNA-binding protein CUGBP1 regulates stability of tumor necrosis factor mRNA in muscle cells: implications for myotonic dystrophy. *J. Biol. Chem.* **283,** 22457–22463 (2008).

33. Ladd, A. N. & Cooper, T. A. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.* **3,** reviews0008.1–reviews0008.16 (2002).

34. Licatalosi, D. *et al.* Mechanisms of alternative mRNA processing in the brain revealed by HITS-CLIP. *Nature* doi:10.1038/nature07488 (this issue).

35. Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature Struct. Mol. Biol.* **12,** 691–698 (2005).

36. Kim, H. H. & Gorospe, M. GU-rich RNA: expanding CUGBP1 function, broadening mRNA turnover. *Mol. Cell* **29,** 151–152 (2008).

37. Wu, J. I., Reed, R. B., Grabowski, P. J. & Artzt, K. Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl Acad. Sci. USA* **99,** 4233–4238 (2002).

38. Paz, R. D. *et al.* Increased expression of activity-dependent genes in cerebellar glutamatergic neurons of patients with schizophrenia. *Am. J. Psychiatry* **163,** 1829–1831 (2006).

39. Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15,** 50–65 (2001).

40. Schroth, G. P., Luo, S. & Khrebtukova, I. Transcriptome analysis using high-throughput DNA sequencing. *Methods Mol. Biol.* (in the press).

41. Illumina, Inc. Transcriptome Analysis: mRNA-Seq. ⟨http://www.illumina.com/pages.ilmn?ID=291⟩ (2008).

42. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human genome by next generation sequencing. *Nature Genet.* (in the press).

LETTERS

Water vapour jets inside the plume of gas leaving Enceladus

C. J. Hansen[1], L. W. Esposito[2], A. I. F. Stewart[2], B. Meinke[2], B. Wallis[1], J. E. Colwell[3], A. R. Hendrix[1], K. Larsen[2], W. Pryor[4] & F. Tian[5]

A plume of water vapour escapes from fissures crossing the south polar region of the Saturnian moon Enceladus[1–6]. Tidal deformation of a thin surface crust above an internal ocean could result in tensile and compressive stresses that would affect the width of the fissures[7]; therefore, the quantity of water vapour released at different locations in Enceladus' eccentric orbit is a crucial measurement of tidal control of venting. Here we report observations of an occultation of a star by the plume on 24 October 2007 that revealed four high-density gas jets superimposed on the background plume. The gas jet positions coincide with those of dust jets reported elsewhere[8] inside the plume. The maximum water column density in the plume is about twice the density reported earlier[2]. The density ratio does not agree with predictions[7]—we should have seen less water than was observed in 2005. The ratio of the jets' bulk vertical velocities to their thermal velocities is 1.5 ± 0.2, which supports the hypothesis that the source of the plume is liquid water, with gas accelerated to supersonic velocity in nozzle-like channels[9].

The Ultraviolet Imaging Spectrograph (UVIS) on board the Cassini spacecraft had observed Enceladus' plume occulting the star γ Orionis in 2005, showing[2] (1) that the plume is localized in the south polar region, (2) that the plume is dominated by water vapour with a line of sight column abundance of $1.5 \times 10^{16} \, cm^{-2}$, and (3) that the flux of water coming from Enceladus ($\sim 150 \, kg \, s^{-1}$) can supply the neutral oxygen pervading the system[10] and maintain Saturn's E ring.

On 24 October 2007, UVIS observed ζ Orionis (Alnitak) occulted by the plume, and we report these results here. UVIS again employed its High Speed Photometer (HSP; wavelength range 1,140–1,900 Å, 2 ms integration) and its Far Ultraviolet Spectrograph (FUV; 1,115–1,914 Å, 1.56 Å resolution, 5 s integration)[11]. Table 1 summarizes the observation geometry. HSP data (Fig. 1, summed to 200 ms) allow us to identify statistically significant events in the profile with a procedure similar to that used for Saturn ring occultations[12,13], described

in Supplementary Information. Assuming a Poissonian distribution, we calculate the value of m, the number of such events one would expect to occur by chance in the entire data set. We identify four events with $m \ll 1$, shown as 1, 2, 3 and 6 in Fig. 1. These indicate local enhancements of the plume outflow that we term 'jets'. We compare them to the dust jets identified by Spitale and Porco[8], which issue from the fissures named Alexandria, Baghdad, Cairo and Damascus.

Table 2 summarizes the locations and m values for the four events deemed to be significant and due to enhanced density jets, re-labelled a, b, c and d. The ground track of the ray from UVIS to the star is shown in Fig. 2. Figure 3 shows the position of the dust jets as seen from the perspective of UVIS, including their angle to the vertical[8]. As some of the dust jets overlap, not all eight are expected to be visible as separate events. Small increases in opacity may correspond to dust
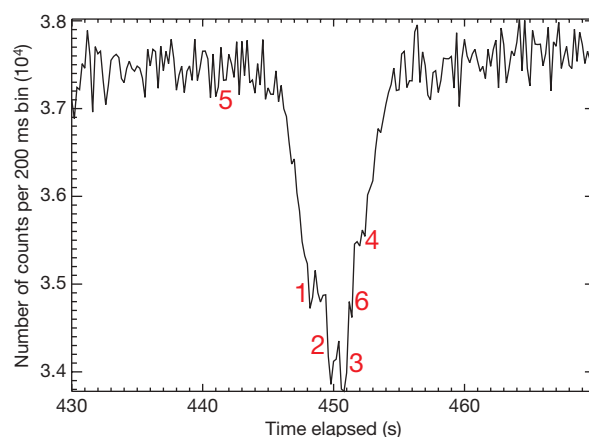
Figure 1 | High Speed Photometer (HSP) data binned to 200 ms intervals. The x axis shows time elapsed in seconds from the beginning of the observation. The broad absorption is due to the water vapour in the large plume that absorbs starlight in the far ultraviolet wavelength range to which the HSP is sensitive: 1,140–1,900 Å. The unoccluded signal of the star is $\sim 3.76 \times 10^4$ counts per 200 ms; attenuation as the starlight passes through the plume reduces the counts to a low of 3.3778×10^4, or about 10%. Smaller features relative to the large absorption are labelled with numbers. Events 2 and 3 have $m < 10^{-4}$, so they are likely to be real features of the plume. Events 1 and 6 have $m < 0.03$. The last two, 4 and 5, have m values of 0.798 and 0.20, respectively, so one might expect such events by chance 80% or 20% of the time. In subsequent plots features 1, 2, 3 and 6 are re-labelled a, b, c and d. 'Plume' is used to describe the large general water vapour flux coming from Enceladus' south pole, seen in this data as the broad absorption feature, and 'jet' refers to the narrow, enhanced density features detected here as absorption increases within the wide feature.

Table 1 | Occultation geometry and timing information

Parameter	Value
Start time of observation (UTC, h:min:s)	24 October 2007 16:59:50.3
Closest point of ray to star to limb (km)	15.64
Time of closest point to limb, maximum attenuation (h:min:s)	17:07:19.8, 17:07:20.0
Spacecraft velocity relative to Enceladus (km s^{-1})	22.57
Spacecraft range to Enceladus (km)	636,442
Subspacecraft latitude/longitude, phase angle (deg)	−2.48/340.0 W, 69.1
True anomaly (deg)	254.13

The ray to the star made a horizontal cut (approximately parallel to the limb of Enceladus) through the plume, rather than a vertical cut as in 2005.

[1]Jet Propulsion Laboratory/California Institute of Technology, Pasadena, California 91109, USA. [2]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80303, USA. [3]Planetary Sciences Group, University of Central Florida, Orlando, Florida 32816, USA. [4]Science Department, Central Arizona College, Coolidge, Arizona 85228, USA. [5]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Table 2 | Locations of water vapour jets

Feature number	Feature letter	m	Occultation latitude (deg)	Occultation west longitude (deg)	Likely associated dust jet
1	a	0.032	−79	82	Cairo V
2	b	0.000008	−86	112	Alexandria IV
3	c	0.00056	−86	192	Baghdad VI
6	d	0.026	−84	217	Damascus II

The feature number in column 1 refers to the statistical analysis that identified six possible events, four of which were judged to be real jets. These four statistically significant events are re-labelled a, b, c and d. In the third column, m is the number of such events one would expect to occur by chance in the data set. The latitudes and longitudes are for the points on Enceladus nearest to the ray from the spacecraft to the star at the time of the absorption feature. The final column is the dust jet[8] we associate with the gas jet.

jets at Baghdad VII, Baghdad I and Cairo VIII, even though they were not deemed statistically significant, and the breadth of the overall occultation is consistent with the spread and tilt of the jets from Baghdad VII and Damascus III.

Typical jet width is 10 km at altitude $z = 15$ km. Because the width of the two most prominent jets is about one-eighth that of the broader plume, their gas density needs to be twice that of the surrounding plume at 15 km altitude to yield the 12% opacity increase seen in them. An alternative explanation is that the excess opacity in the HSP profile could be due to solid micrometre-sized grains lofted by the jets. The mass of ice grains in that case would far exceed the mass of gas in the jet, and it is therefore unlikely that UVIS sees solid ice grains, as discussed in Supplementary Information.

The separation of the two prominent jets allows us to determine the ratio of the vertical bulk velocity, v_b, to the thermal velocity, v_t. As described and shown in Supplementary Fig. 1, the dip in our profile between the features b and c requires the ratio v_b/v_t to be in the range 1.5 ± 0.2. Lower ratios result in the jets merging. Supersonic flow is consistent with a model that treats the sulci (narrow fissures) as

Figure 3 | **Jets as viewed from the spacecraft.** The perspective of this plot shows how jets would appear to UVIS. Plotted here as a function of time are: the altitude above Enceladus' limb of the line-of-sight from Cassini to the star (dashed line); (2) the attenuation of the HSP signal ('opacity'; solid line); and (3) the projections of the eight jets seen by the ISS instrument[8] into the plane of the figure (lines labelled B7 to D3). The jets have each, for purposes of illustration, been assigned a length of 50 km. Here the labels (the letters A, B, C and D) refer to the 'tiger stripes' Alexandria, Baghdad, Cairo and Damascus, respectively, and the original Roman numerals have been replaced by Arabic numerals. Jet B7 was intersected 35 km from its source on Enceladus' surface; the corresponding distances for jets B6 and C8 were within 18 ± 1 km, and for the others, within 23 ± 2 km.

nozzles, channelling gas to the surface[9]. This model predicts the decoupling and slowing of the grains in the dust jets relative to the vapour, explaining the difference in dust particle velocities derived by ISS[6] and gas velocities derived from UVIS data[14].

In the FUV spectrum, the ratio of occulted to unocculted signal, I/I_0, was compared to the predicted absorption using recent water vapour cross-sections[15]. Two 5 s integrations span elapsed time 445–455 s in the observation. The best fits to these two spectra were obtained using line-of-sight column abundances of 1.26×10^{16} cm^{-2} for the first and 1.36×10^{16} cm^{-2} for the second, both with error bars of $\pm 1.4 \times 10^{15}$ cm^{-2}. We have recomputed the flux of water molecules coming from Enceladus, with these column densities and the vertical velocity derived for 2007 (see Supplementary Information), using the same methodology[2] as in 2005. The flux is $\sim 6.5 \times 10^{27}$ molecules s^{-1} or ~ 200 kg s^{-1}.

The HSP attenuation profile (see Fig. 3) gives a maximum H$_2$O slant column abundance of 2.6×10^{16} cm^{-2} at ~ 15 km altitude at the centre of the plume—this peak value is not resolved in the 5 s FUV integration time. The HSP measured 1.5×10^{16} cm^{-2} at 15 km in 2005, showing an increase by a factor of 1.7 from 2005 to 2007.

With two observations acquired at substantially different times, UVIS occultation data test the hypothesis that predicts eruptive activity will vary as a function of where Enceladus is in its eccentric orbit—that is, that fissures will open and close as tidal stresses apply tension and compression to them[7]. Enceladus' true anomalies at the times of the 2005 and 2007 occultations are 98° and 254°, respectively, so that Enceladus was 0.27 of an orbit away from periapsis in 2005 and 0.76 in 2007. A current model predicts the fissures should have been under tension in 2005 (and thus open), while in 2007 the stresses are compressive and the fissures should have been closed[7]. Whether eruptive activity is completely on/off or just modulated, the column density in 2005 should have been higher than in 2007, the opposite of what UVIS observed. A newer model that incorporates physical libration of Enceladus may resolve this discrepancy[16].

Another significant result comes from analysis of this occultation: the composition of the plume reflects geochemistry in Enceladus' interior[17,18]. Carbon monoxide has been proposed as a possible constituent of the plume at a level of ~ 3%, based on data from Cassini's

Figure 2 | **Locations of gas jets detected by UVIS compared to locations of dust jet clusters.** This south polar plot is taken from Spitale and Porco[8] and shows their locations of the dust jets on the Alexandria, Cairo, Baghdad and Damascus sulci (fissures), labelled with roman numerals. The blue line shows the projection of the ray to ζ Orionis mapped onto the surface. The UVIS line-of-sight is perpendicular to the blue line. The attenuation features a, b, c and d are indicated along the line, in addition to the ingress and egress of the starlight through the plume. Small plus signs on the blue line mark 1 s time ticks. Features a, b, c and d correlate to dust jets Cairo V, Alexandria IV, Baghdad VI and Damascus II, respectively.

Ion and Neutral Mass Spectrometer (INMS)[5]. The INMS detects a species with a mass of 28 AMU, which could be CO or N_2. Absorption due to CO was not observed in the UVIS FUV spectrum in either 2005 or 2007. The UVIS would have been able to detect CO: the formal 2σ upper limit is $3.6 \times 10^{14}\,\mathrm{cm^{-2}}$, corresponding to a mixing ratio with H_2O of 3.0%. Our non-detection of CO thus excludes 3% CO in the plume at the 2σ level.

1. Dougherty, M. K. et al. Identification of a dynamic atmosphere at Enceladus with the Cassini Magnetometer. Science 311, 1406–1409 (2006).
2. Hansen, C. J. et al. Enceladus' water vapor plume. Science 311, 1423–1425 (2006).
3. Spahn, F. et al. Cassini dust measurements at Enceladus and implications for the origin of the E ring. Science 311, 1416–1418 (2006).
4. Spencer, J. R. et al. Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. Science 311, 1401–1405 (2006).
5. Waite, J. H. et al. Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure. Science 311, 1419–1422 (2006).
6. Porco, C. C. et al. Cassini observes the active south pole of Enceladus. Science 311, 1393–1401 (2006).
7. Hurford, T. A., Helfenstein, P., Hoppa, B. V., Greenberg, R. & Bills, B. Eruptions arising from tidally controlled periodic openings of rifts on Enceladus. Nature 447, 292–294 (2007).
8. Spitale, J. N. & Porco, C. C. Association of the jets of Enceladus with the warmest regions on its south-polar fractures. Nature 449, 695–697 (2007).
9. Schmidt, J., Brilliantov, N., Spahn, F. & Kempf, S. Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. Nature 451, 685–688 (2008).
10. Esposito, L. W. et al. Ultraviolet imaging spectroscopy shows an active Saturnian system. Science 307, 1251–1255 (2005).
11. Esposito, L. W. et al. The Cassini Ultraviolet Imaging Spectrograph investigation. Space Sci. Rev. 115, 299–361 (2004).
12. Colwell, J. E. et al. Voyager photopolarimeter observations of Uranian ring occultations. Icarus 83, 102–125 (1990).
13. Esposito, L. W., Meinke, B. K., Colwell, J. E., Nicholson, P. D. & Hedman, M. M. Moonlets and clumps in Saturn's F ring. Icarus 194, 278–289 (2008).
14. Tian, F., Stewart, A. I. F., Toon, O. B., Larsen, K. W. & Esposito, L. W. Monte Carlo simulations of the water vapor plumes on Enceladus. Icarus 188, 154–161 (2007).
15. Mota, R. et al. Water VUV electronic state spectroscopy by synchrotron radiation. Chem. Phys. Lett. 416, 152–159 (2005).
16. Hurford, T. A., Bills, B., Greenberg, R., Hoppa, G. V. & Helfenstein, P. How libration affects strike-slip displacement on Enceladus. Lunar Planet. Sci. Conf. XXXIX, abstr. 1826 (2008).
17. Kieffer, S. W. et al. A clathrate reservoir hypothesis for Enceladus' south polar plume. Science 314, 1764–1766 (2006).
18. Matson, D. L., Castillo, J. C., Lunine, J. & Johnson, T. V. Enceladus' plume: Compositional evidence for a hot interior. Icarus 187, 569–573 (2007).

LETTERS

Harnessing optical forces in integrated photonic circuits

Mo Li[1,2], W. H. P. Pernice[1,2], C. Xiong[1,2], T. Baehr-Jones[3], M. Hochberg[3] & H. X. Tang[1,2]

The force exerted by photons is of fundamental importance in light–matter interactions. For example, in free space, optical tweezers have been widely used to manipulate atoms and microscale dielectric particles[1,2]. This optical force is expected to be greatly enhanced in integrated photonic circuits in which light is highly concentrated at the nanoscale[3,4]. Harnessing the optical force on a semiconductor chip will allow solid state devices, such as electromechanical systems, to operate under new physical principles. Indeed, recent experiments have elucidated the radiation forces of light in high-finesse optical microcavities[5–7], but the large footprint of these devices ultimately prevents scaling down to nanoscale dimensions. Recent theoretical work has predicted that a transverse optical force can be generated and used directly for electromechanical actuation without the need for a high-finesse cavity[3]. However, on-chip exploitation of this force has been a significant challenge, primarily owing to the lack of efficient nanoscale mechanical transducers in the photonics domain. Here we report the direct detection and exploitation of transverse optical forces in an integrated silicon photonic circuit through an embedded nanomechanical resonator. The nanomechanical device, a free-standing waveguide, is driven by the optical force and read out through evanescent coupling of the guided light to the dielectric substrate. This new optical force enables all-optical operation of nanomechanical systems on a CMOS (complementary metal-oxide-semiconductor)-compatible platform, with substantial bandwidth and design flexibility compared to conventional electrical-based schemes.

The forces of light stem from two major mechanisms: radiation pressure and the transverse gradient force. Forces induced by radiation pressure have been extensively studied in high-finesse optical cavities, in which the momentum of light is transferred to the mirrors forming the cavity and therefore results in an axial force. Similar radiation pressure was also found in high-finesse micro-toroids[5] and microsphere[8] optical resonators. The transverse gradient force, on the other hand, originates from the lateral gradient of a propagating light field. Recently, it was theoretically predicted that this seemingly small force could be significant in submicrometre-scale photonic waveguides, because the gradient of the light is enhanced by orders of magnitude[3]. This transverse force does not require a reflective surface; thus it is more versatile for operation in planar structures than the radiation pressure force.

In this work, we demonstrate the direct detection of the optical force in an integrated silicon photonic circuit, and establish that this force can be used to drive nanoscale mechanical devices. We use a simple, yet very generic, configuration to demonstrate the optical force effect (Fig. 1). The force on the free-standing waveguide arises when the guided light evanescently couples to the dielectric substrate. This new force generation mechanism only requires one single-mode waveguide, as opposed to the two coupled waveguides proposed in

recent theoretical work[3,4]. The force can be calculated by integrating the Maxwell stress tensor over the surfaces of the waveguide. In an alternative picture, the eigenstate of the propagating optical mode is shifted as the light enters the free-standing waveguide. Thus a corresponding force arises from the adiabatic change of the energy of the eigenmode[3], as detailed in Supplementary Information. Our main results from both formalisms are presented in Fig. 1. The asymmetric guided mode resulting from coupling to the substrate (Fig. 1b) yields the field gradient required to produce a net optical force. At a given optical input power, the optical force on the waveguide (Fig. 1d) as well as the effective refractive index (Fig. 1c) strongly depends on the separation between waveguide and substrate. The calculations reveal that as this separation is reduced from 500 nm to 50 nm, the magnitude of the optical force increases from 0.1 to 8 pN μm^{-1} mW^{-1}.

To experimentally demonstrate this optical force, we exploit the high sensitivity offered by nanoelectromechanical systems (NEMS)[9,10], integrated in a silicon photonic circuit that was fabricated with a CMOS-compatible process (Fig. 2). Its simplest form consists of a single mode strip waveguide between two grating couplers[11] (Fig. 2a). A portion of the waveguide is released from the substrate by wet etching. The resulting NEMS beam is mechanically supported by two specially designed, low loss multimode interference (MMI) structures (Fig. 2c). The separation between the beam and the substrate (300–600 nm) can be tuned by varying the etching time. A scanning electron microscope image of a typical suspended device (10 μm long, 500 nm wide, 110 nm thick) is shown in Fig. 2b.

We apply two tunable diode lasers in a pump–probe scheme to quantify the optical force (Fig. 2a). The actuation laser is amplitude modulated to generate a time-varying force on the beam. The second laser acts as the probe for displacement readout. In Fig. 3a, the transmission spectrum of the device from Fig. 2b is presented. The peak transmission of the system is 4×10^{-3} (or -24 dB), with most of the optical loss occurring at the grating couplers (-10 ± 1 dB each) and the two MMI couplers (-4 dB total). The envelope results from the transmission pass-band of the grating coupler with a width of ~20 nm. The transmission oscillates with a period of 1.9 nm due to the Fabry–Perot interferometer formed by the two grating couplers. These fringes enable interferometric detection of the motion of the device, using the probe laser at the maximum slope of the transmission (the magenta marker in Fig. 3a). This point corresponds to maximum detection responsivity. The wavelength of the actuation laser is fixed at the peak of the transmission (the blue marker in Fig. 3a).

When the actuation laser is turned off, we are able to observe the thermomechanical ('Brownian') motion of the beam (Fig. 3c, d). The resulting displacement noise power spectral density (PSD) provides a reliable way to calibrate the sensitivity of the system[12]. As the sensitivity depends on the probe laser power and the interferometer

[1]Department of Electrical Engineering, [2]Department of Mechanical Engineering, Yale University, New Haven, Connecticut 06511, USA. [3]Department of Electrical Engineering, University of Washington, Seattle, Washington 98195, USA.
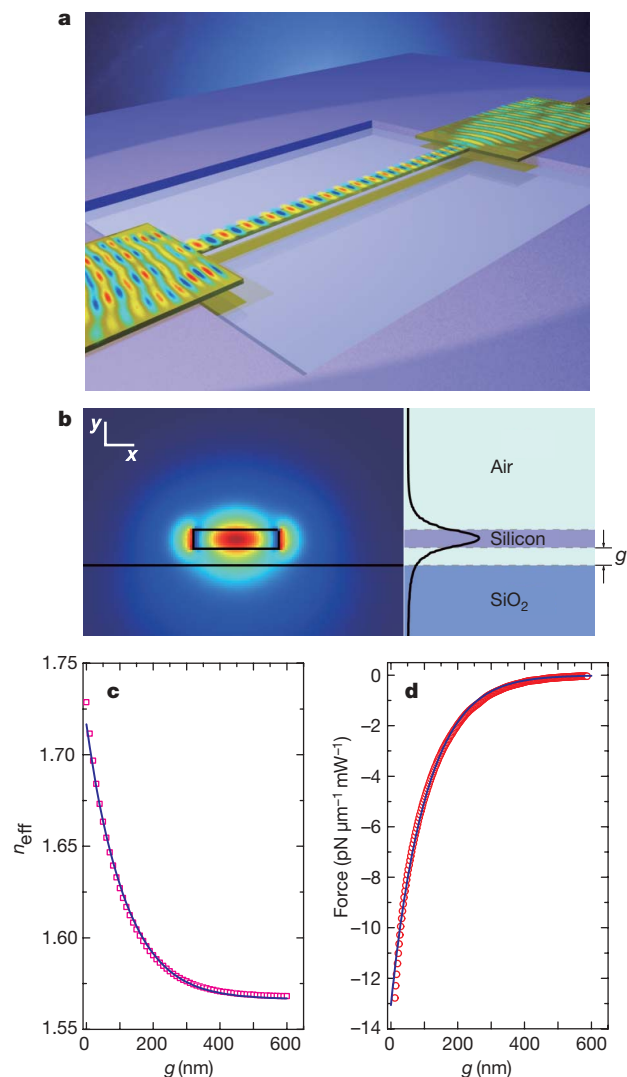
Figure 1 | **The substrate coupled waveguide gradient force. a,** Three-dimensional schematic illustration of a free-standing waveguide beam supported by two multimode interference structures, with an overlay of the optical mode plot. **b,** Left, finite element simulation result showing the E_x component of the optical fields in the waveguide evanescently coupled to the dielectric substrate at the separation gap of size g. Right, the black curve shows the magnitude of the E_x at the cross-section through the centre of the waveguide. We note that E_x is continuous at the bottom surface of the waveguide and the top surface of the substrate. This is different from the situation of side-coupled waveguides. **c, d,** The effective refractive index of the waveguide (**c**) and the optical force on the waveguide (**d**) strongly depend on the separation g. The solid line is the fitting to the model.

Figure 2 | **Schematics of the experimental set-up and device system. a,** The measurement set-up. **b,** Scanning electron micrograph of a free-standing, 10-μm-long NEMS beam. **c,** The finite-difference time-domain simulation result showing the mode conversion through the MMI coupler into the NEMS beam (colour showing the H_y component of the optical mode). **d,** Finite element simulation showing the mechanical mode shape of the waveguide beam, with the strain distribution displayed in colour.

transmission, a separate calibration procedure at fixed temperature is performed for each measured device. The PSD of the beam at resonance can be expressed as $S_z^{1/2} = \sqrt{4 k_B T Q/(k \omega_0)}$, where k_B is the Boltzmann constant, T is the absolute temperature (300 K), Q is the mechanical quality factor, ω_0 is the angular resonance frequency and k the spring constant. Using standard beam theory as well as finite element simulations, the spring constant k of the beam can be determined. Hence the displacement responsivity (transduction gain) of the measurement system can be calibrated. Figure 3c shows the measured noise spectrum of a 10-μm-long beam with resonance frequency of 8.87 MHz and quality factor of 1,850 in vacuum, using 15 μW probe light power on the beam (determined by the procedure described in Methods). The noise floor corresponds to a displacement noise PSD of 7.2×10^{-14} m Hz$^{-1/2}$. Using 150 μW probe light power on a 13-μm-long device, we achieved improved displacement

sensitivity of 1.8×10^{-14} m Hz$^{-1/2}$ (Fig. 3d). This notable sensitivity is achieved without the aid of a high-finesse cavity at room temperature. For devices of similar dimensions, equivalent sensitivity has only been attained with millikelvin techniques[10,13].

When the actuation laser is on and modulated at various amplitudes, we measure the dynamic response of the device with the probe laser (Fig. 3b). The amplitude of the optical power on the beam is derived from the transmitted power, taking into account the power enhancement by the intra-cavity effect of the Fabry–Perot interferometer[14]. The corresponding driving force on the beam is calculated by $F = k\Delta z(\omega_0)/Q$, where $\Delta z(\omega_0)$ is the beam's resonance amplitude as calibrated by the aforementioned thermomechanical noise measurement procedure. For the beam measured in Fig. 3b with a substrate separation of 360 nm, the optical force normalized to the beam length and the optical power is 0.5 ± 0.1 pN μm^{-1} mW^{-1}. This value is in line with our modelling (Fig. 1d) and theoretical values presented in the literature[3].

Alternatively, the nonlinear response of the beam's mechanical resonance provides an independent way to assess the optical force. As shown in Fig. 3b, the vibration amplitude of the beam increases linearly with the actuation power until the critical amplitude is reached. Afterwards the vibration behaves like a nonlinear Duffing oscillator[15]. The critical amplitude, experimentally determined from the 'backbone' curve in Fig. 3b, is 2.5 nm, very close to the theoretical value of 2.2 nm expected for a nonlinear beam resonator[15].

The optical force can be tuned by varying the separation between the beam and the substrate (see Fig. 1d). To confirm this theoretical prediction, a series of devices with varying separation distances were measured (Fig. 3e). The measured optical force shows a clear dependence on the separation size—increasing with decreasing separation size, indicating the same trend as the theory. Discrepancies between the theory and the experiment are expected, because the simple model does not account for the varying gap under the beam clamping points.

It is also known that photothermal effects can cause actuation of mechanical oscillations through optical absorption of visible light in composite structures[16–18]. Our devices are made of single-crystal silicon, which has very low adsorption at infrared wavelengths, so the photothermal force is expected to be weak[19]. To provide unambiguous proof of the observed optical force, it is important to fully separate photothermal effects and optical forces. We differentiate these effects by studying the device's dynamic response over a wide frequency range, as the photothermal effect has a characteristic time constant but the optical force is wideband[16,19].

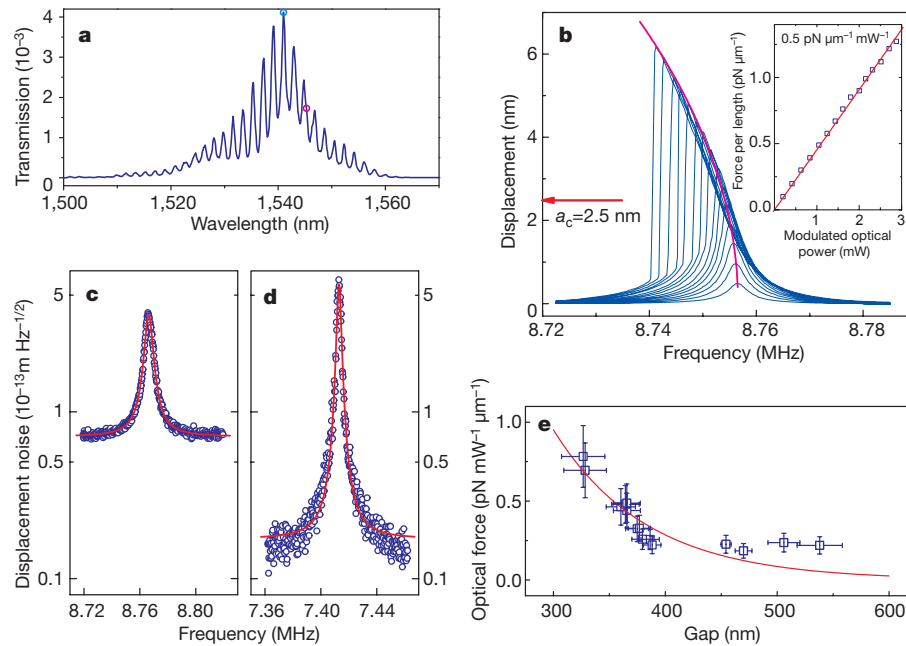**Figure 3 | Device characterization and experimental demonstration of the waveguide gradient force. a,** The typical transmission spectrum of the devices, showing interference fringes of the internal Fabry–Perot interferometer formed by the input and output couplers. The wavelengths of the actuation and probe light are marked with the blue and magenta markers, respectively. **b,** Resonance response curves of a 10-μm-long waveguide beam at varying modulation levels of the actuation light. When the vibration amplitude exceeds the critical amplitude a_c, the response shows a strong softening nonlinearity. The critical amplitude of 2.5 ± 0.1 nm, determined from the backbone curve (magenta line), agrees well with the

theoretical value of 2.2 nm. Inset, the vibration amplitude versus modulated optical power on the device shows a linear response. Fitting the data gives the optical force of 0.5 ± 0.1 pN μm^{-1} mW^{-1}. **c, d,** Measured noise PSD, showing Lorentzian thermomechanical resonance peaks of waveguide beams 10 μm long (**c**) and 13 μm long (**d**). The best minimum detection level achieved shown in **d** is ∼18 fm Hz$^{-1/2}$. **e,** The optical force is measured on devices with various beam lengths and substrate separation sizes. The measurement (blue open squares) with error bars (horizontal: s.d., $n = 5$ measurements; vertical: s.d., $n = 4$ measurements) and the theoretical model (red line) show qualitative agreement.

To assess the thermal effect, we designed devices in Mach–Zehnder interferometer (MZI) configuration, in which the two arms are identical except for a $\Delta L = 100$ μm path difference (Fig. 4a). The upper arm contains a MMI supported nanomechanical beam; the reference arm contains a dummy unsuspended beam. The transmission spectrum

with the beam released shows the interference fringes of a well-matched MZI with ∼30 dB extinction ratio (Fig. 4b). The MZI allows us to measure the refractive index change $\delta n(f)$ of the device as a function of the modulation frequency of the pump (see Methods). Comparison of the response measured on the MZI device before
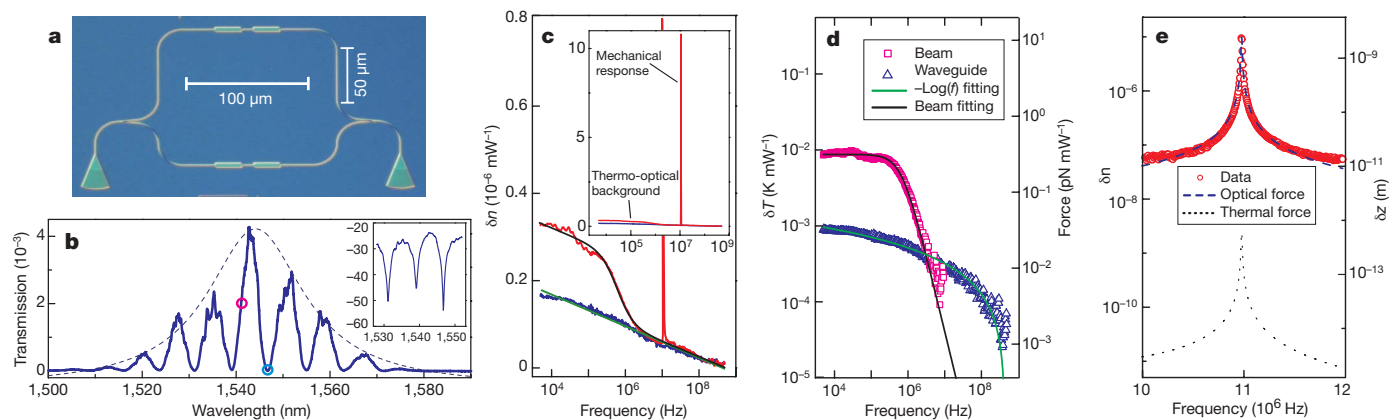
Figure 4 | Measurement of the thermal response of the device. a, Optical micrograph of the Mach–Zehnder interferometer (MZI) device. The path length of the top arm is 100 μm longer than that of the bottom arm. **b,** Transmission spectrum of the MZI device, showing the interference fringes with nice visibility even after release. The wavelengths of the actuation and detection lasers are marked with the blue and magenta markers, respectively. The inset shows an extinction ratio of ∼30 dB (vertical axis, transmission in dB; horizontal axis, as main plot). **c,** Wide frequency range measurement of the effective index response of the MZI device with (red) and without (blue) released NEMS beam. Blue trace has a $-\log(f)$ dependence on the frequency (line, green) due to the thermo-optical nonlinear effect. The distinct response shown by the red trace at low

frequency is due to the slow thermal response of the suspended beam. Inset, expanded plot showing the relative height of the resonance peak with respect to the background (axes as main plot). **d,** Dynamic temperature variation of the released NEMS beam versus frequency (open squares, magenta) with the fitting of the first order thermal model (line, black), as well as the unreleased waveguide (open triangles, blue) with $-\log(f)$ fitting (line, green). The corresponding thermal force is marked on the right axis. **e,** The measured mechanical resonance response of the beam (open circles, red). The contribution from the optical force (dashed line, blue) is three orders of magnitude higher than that from the thermal force (dotted line, black). The corresponding beam displacement is marked on the right axis.

(blue) and after (red) the release (Fig. 4c) reveals the thermal response of the nanomechanical beam.

The response of the unreleased MZI's average refractive index change, δn_w, shows a clear $-\log(f)$ dependence, which is the characteristic response of diffusive heating from a surface heater on a semi-infinite medium[20]. We note that unlike in silica[19], the measured response is dominated by the thermal nonlinear effect of silicon, which overshadows other effects such as the DC Kerr effect[21] and two-photon adsorption[22]. Thus, knowing the thermo-optical coefficient of silicon ($\partial n/\partial T = 1.86 \times 10^{-4}\,\mathrm{K}^{-1}$), we can extract the averaged dynamic temperature variation of the substrate supported waveguide from the measured effective index change using $\delta T = \delta n/(\partial n/\partial T)$, as shown in Fig. 4d (blue). Furthermore, the slope of this curve, given by $-\alpha P \ln(10)/2\pi\kappa$, is commonly used in electrical domain '3ω methods' to measure the thermal conductivity κ of a material[20]. Here it allows us to evaluate the thermal adsorption rate α of the waveguide at propagating power P. We find that the α value in our device is $0.04\,\mathrm{cm}^{-1}$ ($0.2\,\mathrm{dB\,cm}^{-1}$), which accounts for about 3% of the total waveguide propagation loss ($1.2\,\mathrm{cm}^{-1}$ or $5\,\mathrm{dB\,cm}^{-1}$). (See Supplementary Information for a detailed analysis.)

In the released MZI, the measured response is composed of three parts $\delta n = [0.7\,\Delta L\,\delta n_w + 0.63\,l\,(\delta n_m + \delta n_b)]/(\Delta L + l)$, with δn_w as discussed above, δn_b the average refractive index change due to the temperature rise in the suspended NEMS beam of length l, and δn_m the mechanical resonance of the NEMS beam (the narrow peak at mechanical resonance $\sim 11\,\mathrm{MHz}$). The numerical factors compensate for the different power on the NEMS beam and the waveguide due to the insertion loss of two MMI couplers ($-4\,\mathrm{dB}$). Thus, the thermal response of the nanomechanical beam $\delta n_b(f)$ (Fig. 4d, magenta) can be determined from the difference of the two curves in Fig. 4c. The result shows that the temperature variation of the beam is $\sim 10\,\mathrm{mK\,mW}^{-1}$ at low frequency. As in a typical thermal system (black curve, Fig. 4d), the temperature variation rolls off with $-20\,\mathrm{dB}$ per frequency decade after the characteristic frequency of $\sim 600\,\mathrm{kHz}$, reducing to $\sim 40\,\mathrm{\mu K\,mW}^{-1}$ at the resonant frequency of the beam ($11\,\mathrm{MHz}$). This temperature variation induces differential expansion between the waveguide and the substrate and generates an effective thermal force. We analysed the thermal-mechanical response of the NEMS beam both analytically and numerically (see Supplementary Information). We find that the centre of the beam undergoes a vertical displacement of about 11 pm for every 1 K temperature change on the beam, which is consistent with values in ref. 17. Given the force constant ($\sim 3.6\,\mathrm{N\,m}^{-1}$) of the 10-μm-long device, this thermal response corresponds to a force of $40\,\mathrm{pN\,K}^{-1}$. Normalized to power, the equivalent photothermal force on the device is $0.4\,\mathrm{pN\,mW}^{-1}$ at low frequency, but decreases quickly to only $1.6\,\mathrm{fN\,mW}^{-1}$ at $11\,\mathrm{MHz}$. The displacement $\Delta z(\omega_0)$ of the beam at resonance, as calibrated from the measured resonance peak signal (Fig. 4e), is $2.5\,\mathrm{nm}$ at $1\,\mathrm{mW}$ actuation amplitude. This corresponds to an actuation force of $k\Delta z(\omega_0)/Q = 4.8\,\mathrm{pN\,mW}^{-1}$ on the 10-μm-long beam, as expected from the optical force, which is three orders of magnitude larger than what can be provided by the thermal effect (Fig. 4e). Thus, we conclude that the observed optical force is not due to the thermal effect, which has much lower magnitude and slower frequency response.

The prominent mechanical response of our devices demonstrates that the actuation efficiency of the transverse optical force is as efficient as other types of electrical-based actuation forces commonly used in nanomechanical devices, such as capacitive, piezoelectric and electrothermal forces. This optical force could be further enhanced by coupling the nanomechanical devices with high-finesse cavities or slow-light structures. The all-optical scheme presented here also significantly improves the signal quality by avoiding the parasitic crosstalk commonly encountered in electrical systems. Furthermore, the optical waveguide force offers an enormous operation bandwidth, in practice only limited by the bandwidth of our input–output couplers. This improvement of actuation bandwidth allows us to explore nanomechanical motion

beyond what has been obtained with today's electrical actuators[23–25]. In conjunction with the high transduction gain and low readout noise, this new scheme will enable a broad range of experiments, such as ultrafast detection and stroboscopic measurement.

The system presented here seamlessly integrates nanomechanical devices that are size-matched to photonic waveguides, while offering excellent engineering flexibility. Arrays of nanomechanical resonators can be fabricated along a photonic 'bus' for efficient synchronization and high speed optical interconnection. This way, one can achieve long-range coherent signal processing without locally addressing individual devices. This will eventually lead to large-scale integration of nanomechanical systems and enable new device functions, such as all-optical switching[26], optomechanical signal processing, radio-frequency photonics, and tunable nanophotonics[27].

METHODS SUMMARY

Our integrated photonic devices are fabricated by standard electron beam lithography and dry etching processes on silicon-on-insulator wafers with a silicon layer thickness of 110 nm. The measurements are conducted in a vacuum chamber with a base pressure of $10^{-6}\,\mathrm{torr}$ at room temperature. An array of cleaved fibres aligned to on-chip grating couplers is used to couple the optical signal into and out of the devices. By measuring the total optical transmission of many identical devices, the optical power on the devices is derived. The variation of the transmission is included as a systematic error.

The thermomechanical noise spectrum of the device in the opto-electrical signal domain is used to calibrate the transduction gain of the system. Then the corresponding optical force of the device's driven response is determined using the elastic properties of the nanomechanical resonator.

The thermal effect of the silicon waveguides is studied by measuring the dynamic response to the optical modulation over a wide frequency range. Using the known thermo-optic coefficient of silicon, the temperature variation of the waveguide, as well as the thermal adsorption coefficient, are determined. Thus, the corresponding thermal actuation force can be evaluated from the analytical and numerical models.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Ashkin, A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett.* **24**, 156–159 (1970).
2. Chu, S. Laser manipulation of atoms and particles. *Science* **253**, 861–866 (1991).
3. Povinelli, M. L. *et al.* Evanescent-wave bonding between optical waveguides. *Opt. Lett.* **30**, 3042–3044 (2005).
4. Rakich, P. T., Popovic, M. A., Soljacic, M. & Ippen, E. P. Trapping, corralling and spectral bonding of optical resonances through optically induced potentials. *Nature Photon.* **1**, 658–665 (2007).
5. Kippenberg, T. J., Rokhsari, H., Carmon, T., Scherer, A. & Vahala, K. J. Analysis of radiation-pressure induced mechanical oscillation of an optical microcavity. *Phys. Rev. Lett.* **95**, 033901 (2005).
6. Kippenberg, T. J. & Vahala, K. J. Cavity opto-mechanics. *Opt. Express* **15**, 17172–17205 (2007).
7. Eichenfield, M., Michael, C. P., Perahia, R. & Painter, O. Actuation of micro-optomechanical systems via cavity-enhanced optical dipole forces. *Nature Photon.* **1**, 416–422 (2007).
8. Carmon, T. & Vahala, K. J. Modal spectroscopy of optoexcited vibrations of a micron-scale on-chip resonator at greater than 1 GHz frequency. *Phys. Rev. Lett.* **98**, 123901 (2007).
9. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).
10. Regal, C. A., Teufel, J. D. & Lehnert, K. W. Measuring nanomechanical motion with a microwave cavity interferometer. *Nature Phys.* **4**, 555–560 (2008).
11. Taillaert, D., Bienstman, P. & Baets, R. Compact efficient broadband grating coupler for silicon-on-insulator waveguides. *Opt. Lett.* **29**, 2749–2751 (2004).
12. Sader, J. E., Larson, I., Mulvaney, P. & White, L. R. Method for the calibration of atomic-force microscope cantilevers. *Rev. Sci. Instrum.* **66**, 3789–3798 (1995).
13. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
14. Garmire, E. Criteria for optical bistability in a lossy saturating Fabry-Perot. *IEEE J. Quant. Electron.* **25**, 289–295 (1989).
15. Nayfeh, A. H. & Mook, D. T. *Nonlinear Oscillations* (Wiley, 1979).
16. Kleckner, D. & Bouwmeester, D. Sub-kelvin optical cooling of a micromechanical resonator. *Nature* **444**, 75–78 (2006).
17. Ilic, B., Krylov, S., Aubin, K., Reichenbach, R. & Craighead, H. G. Optical excitation of nanoelectromechanical oscillators. *Appl. Phys. Lett.* **86**, 193114 (2005).

18. Metzger, C. H. & Karrai, K. Cavity cooling of a microlever. *Nature* **432**, 1002–1005 (2004).
19. Schliesser, A., Del'Haye, P., Nooshi, N., Vahala, K. J. & Kippenberg, T. J. Radiation pressure cooling of a micromechanical oscillator using dynamical backaction. *Phys. Rev. Lett.* **97**, 243905 (2006).
20. Cahill, D. G. Thermal-conductivity measurement from 30-K to 750-K – the 3ω method. *Rev. Sci. Instrum.* **61**, 802–808 (1990).
21. Soref, R. A. & Bennett, B. R. Electrooptical effects in silicon. *IEEE J. Quant. Electron.* **23**, 123–129 (1987).
22. Almeida, V. R. & Lipson, M. Optical bistability on a silicon chip. *Opt. Lett.* **29**, 2387–2389 (2004).
23. Huang, X. M. H., Zorman, C. A., Mehregany, M. & Roukes, M. L. Nanodevice motion at microwave frequencies. *Nature* **421**, 496 (2003).
24. Li, M., Tang, H. X. & Roukes, M. L. Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications. *Nature Nanotechnol.* **2**, 114–120 (2007).
25. Masmanidis, S. C. *et al.* Multifunctional nanomechanical systems via tunably coupled piezoelectric actuation. *Science* **317**, 780–783 (2007).
26. Almeida, V. R., Barrios, C. A., Panepucci, R. R. & Lipson, M. All-optical control of light on a silicon chip. *Nature* **431**, 1081–1084 (2004).
27. Huang, M. C. Y., Zhou, Y. & Chang-Hasnain, C. J. A nanoelectromechanical tunable laser. *Nature Photon.* **2**, 180–184 (2008).
28. Almeida, V. R., Panepucci, R. R. & Lipson, M. Nanotaper for compact mode conversion. *Opt. Lett.* **28**, 1302–1304 (2003).

Author Contributions H.X.T. planned and supervised the project; M. L. and H.X.T. conceived the experiment; M.L. fabricated the devices, and did the measurements; W.H.P.P. conducted the force calculation and simulation; C.X. helped with the automation of the pattern generation; M.L., W.H.P.P. and H.X.T. analysed the data and wrote the manuscript; and T.B.-J. and M.H. provided the layout of the couplers and waveguide, and assisted with fabrication of devices.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.X.T. (hong.tang@yale.edu).

METHODS

Device fabrication. Our devices are fabricated on silicon-on-insulator (SOI) wafers manufactured by Soitec (Smart-cut). The silicon layer is thinned down to 110 nm by thermal oxidation and subsequent wet etch. The thickness of the buried oxide layer is 3 μm. The photonic devices (waveguide and couplers) are patterned by electron beam lithography (Vistec VB-6 at Cornell Nanofabrication Facility) and plasma etching down to the oxide layer. The etch release windows are subsequently defined by photolithography and the nanomechanical beams are released from the substrate by wet etching using buffered oxide etchant. To fully suspend the beam from the substrate, the etching time has to be sufficient to form a well-defined gap between the beam and the substrate. This limits the minimum achievable separation size to roughly 300 nm, though in principle, this limitation can be overcome by fabricating side-coupled devices[3].

The waveguide TE mode is selected by polarization controllers and launched into the device through a single-mode polarization-maintaining fibre aligned with the input coupler. Another fibre aligned with the output coupler collects the transmitted signal. The use of grating couplers as optical input/output ports provides a convenient and high throughput interface to large arrays of planar photonic devices using an array of fibres. Coupler optimization is essential in determining the working wavelength range of the photonic circuits and achieving best signal to noise ratio in the measurement. We routinely achieve a coupling efficiency of 10% (or −10 dB) with our fabrication processes. However, it has been shown theoretically that nanoscale silicon waveguides can be coupled to optical fibres with less than 1 dB of loss per coupler with various schemes. Experimental results of 3 dB loss per coupler have been obtained for both edge and grating couplers[11,28].

Calibration of optical power on the device. The optical power on the device is derived from the optical input power P_{in} and the transmitted output power P_{out}. The overall transmission of the device is $T_0 = P_{out}/P_{in} = T_1 T_2$, where T_1 and T_2 denote the transmission before and after the NEMS beam including the grating coupler, the waveguide and the MMI coupler. The typical total insertion loss of the devices is about −24 dB, with contributions from the input and output grating couplers (about −10 ± 1 dB each) and the two MMI couplers (about −4 dB total). The silicon waveguide propagation loss is estimated to be ~5 dB cm^{-1} and its contribution to the total insertion loss (~0.1 dB) can be neglected. Since the device is fabricated in a symmetric fashion—identical grating couplers and MMI couplers on each side of the NEMS beam—T_1 and T_2 are approximately equal, so that $T_1 = T_2 = \sqrt{T_0}$ (the uncertainty of this assumption is discussed in Supplementary Information). Therefore, the optical power delivered to the device can be derived as: $P_{beam} = P_{in} T_1 = P_{out}/T_2 = P_{in}\sqrt{T_0} = P_{out}/\sqrt{T_0}$.

A power enhancement factor G due to the low-finesse Fabry–Perot interferometer formed by the two grating couplers is also taken into account. Fitting the interference fringes of the transmission spectrum with the Airy function enveloped by the grating coupler transmission gives a coefficient of finesse $F = 4R_e/(1 - R_e)^2$, where the effective reflectivity $R_e = R_c T_{MMI}$ takes account of the internal reflectivity of the grating coupler R_c and transmission of the MMI coupler T_{MMI}. When the laser wavelength is at the maximum of the interference fringes, the optical power inside the cavity is enhanced by a factor $G = (1 + R_e)/(1 - R_e)$ (ref. 14). From the measured transmission spectrum in Fig. 3a, we obtain a coefficient of finesse F of 2.0, thus the enhancement factor G is 1.7. Hence the optical power on the device is $P_{beam} = P_{out} G/\sqrt{T_0}$. In the case of a MZ interferometer, the power on the beam is reduced by half, therefore $G = 0.5$.

Calibration of displacement sensitivity and optical force. The displacement sensitivity is calibrated by measuring the device's thermomechanical noise spectrum at the same conditions as in the driven measurement. Since this sensitivity depends on the optical power, the detection laser wavelength and the Fabry–Perot cavity transmission spectrum of different devices, a separate calibration procedure is performed for each device (for example, each data point in Fig. 3e). The measured electrical voltage noise PSD, $S_N(\omega)$ in V^2 Hz^{-1}, of the amplified photodetector signal is measured near the resonance frequency of the beam, showing a peak corresponding to the thermomechanical motion of the NEMS resonator (for example, Fig. 3c and d). The peak is fitted with a Lorentz function $S_v(\omega)$ with a constant baseline: $S_N(\omega) = S_v(\omega) + S_0$. The amplitude of the peak corresponds to the thermomechanical noise spectrum density of the NEMS resonator at resonance: $S_z(\omega_0) = 4k_B TQ/\omega_0 k$ (in m^2 Hz^{-1}). Using standard beam theory as well as finite element simulations, both the resonance frequency ω_0 and the spring constant k of the beam can be determined. The accuracy of the theoretical results can be examined by comparing the theoretical resonance frequency with the measured value and good agreement can regularly be

obtained. Hence the displacement responsivity (transduction gain) of the measurement system can be calibrated to be: $R_v^z = \sqrt{S_z(\omega_0)/S_v(\omega_0)}$ (in m V^{-1}). Furthermore, the displacement sensitivity of the measurement system, that is, the off-resonance output voltage noise floor transferred to the mechanical displacement domain, can be evaluated using R_v^z: $x_n = \sqrt{S_z} \simeq S_0^{1/2}/R_v^z$ (in m Hz$^{-1/2}$). When conducting the driven measurement of a device, the wavelength and power of the probe laser is kept at the calibrated conditions. Thus the driven vibration displacement amplitude of the device $x(\omega)$ can be determined by transferring the measured voltage signal amplitude $v(\omega)$ to the displacement domain using the calibrated R_v^z. Then the driving force applied to the device at the resonance frequency is determined as $F = kx(\omega_0)/Q$.

The calibrated displacement responsivity can also be tested by comparing the measured nonlinear critical amplitude with its theoretical value. The critical amplitude can be determined by fitting the backbone curve which connects the maxima of the resonance response curves using $f_{peak} = f_0(1 - 1/\sqrt{3}Q(a_{peak}/a_c)^2)$. Its theoretical value is given by $a_c = \omega_0 (l/\pi)^2 (\sqrt{3}\rho/EQ)^{1/2}$ (l is the beam length, ρ is the beam material density, E is the material Young's modulus)[15].

Derivation of the temperature response. Silicon is known to have a strong thermal nonlinear effect with a thermo-optic coefficient of $\alpha_{TO} = \partial n/\partial T = 1.86 \times 10^{-4}$ K^{-1}, which dominates other nonlinear optical effects at low frequency. We use an MZI configuration to measure the thermal effect of the photonic device. The path length of the top arm of the MZI is 100 μm longer than that of the bottom arm. Identical NEMS beams and MMI couplers are included in the top and bottom arms. Thus when none of the NEMS beams are suspended from the substrate, the system is used in a matched interferometer configuration. Using the pump–probe method described in the main text, the thermal response of the unbalanced 100 μm silicon waveguide in the top arm to the modulated pump light is measured. From the measured response, we can derive the dynamic temperature variation of the waveguide.

The transmission of the MZI can be expressed as: $T(\phi) = T_0[1 + \cos(\phi)]/2$. Here $\phi = 2\pi n\Delta L/\lambda$ is the relative phase difference between the two arms, n is the effective refractive index of the silicon waveguide, $\Delta L = 100$ μm is the path difference, λ is the optical wavelength and T_0 denotes the maximum transmission when the phase difference equals $\phi = 2N\pi$, N is an integer.

During the measurement, the detection light wavelength λ_d is fixed at a transmission of $0.5T_0$ (quadrature point) and hence the phase difference is $\phi = (2N + 1.5)\pi$. The variation of the transmission depends approximately linearly on the variation of ϕ. Using $T(\varphi) = T_0 [1 + \cos(\varphi)]/2 = T_0 [1 + \cos((2N + 3/2)\pi + \delta\varphi)]/2 = T_0 [1 + \sin(\delta\varphi)]/2$ we then find $\delta T(\varphi) \simeq T_0 \delta\varphi/2 = T_0(\pi \Delta L/\lambda_d) \delta n$.

In silicon devices, the thermo-optic effect dominates the measured dynamic response so that the corresponding variation of the effective index is given as $\delta n = \alpha_{TO}\delta T$, where δT is the temperature variation of the waveguide due to the pump laser modulation. The measured dynamic transmitted detection laser power δP_d^{out} through the device is:

$$\delta P_d^{out} = \delta T(\varphi)P_d^{in} = (T_0 \cdot P_d^{in})(\pi \Delta L/\lambda_d)\delta n = 2P_d^{out}(\pi \Delta L/\lambda_d)\delta n$$

Here P_d^{in} and P_d^{out} are the input and output DC power of the detection laser, respectively. Thus the corresponding variation of the effective index can be derived from the measured detection laser signal as: $\delta n = (\delta P_d^{out}/P_d^{out})(\lambda_d/2\pi \Delta L)$. Here $\delta P_d^{out}/P_d^{out}$ equals the ratio of the photodetector's AC to DC signal: v_d^{AC}/v_d^{DC}, which is readily measured. Then the corresponding effective temperature variation of the waveguide is calculated as $\delta T = \delta n/\alpha_{TO}$. The measured value is then normalized to the modulation amplitude of the pump laser power on the device as given by $P_p^M = P_p^0 V_M/V_\pi$, where P_p^0 is the static power of the pump laser on the device, V_M is the applied modulation voltage on the electro-optic modulator and V_π is the full modulation voltage of the modulator. When the response of the waveguide and the NEMS beam are compared, the difference of the optical power level before and after the MMI couplers (insertion loss ~2 dB each) has to be considered. The averaged response of the released MZI device is given by $\delta n = [\gamma_1 \Delta L\delta n_w + \gamma_2 l (\delta n_m + l\delta n_b)]/(\Delta L + l)$, where the factor $\gamma_1 = (1 + T_{MMI}^2)/2 \approx 0.7$ and $\gamma_2 = T_{MMI} \approx 0.63$. First the response of the device is measured over a wide frequency range with a lock-in amplifier from 5 kHz to 100 kHz (EG&G 5120) and a network analyser from 100 kHz to 500 MHz (HP 4195A); then the corresponding effective index and temperature variation response is calculated with the procedure described above.

LETTERS

Total synthesis of bryostatin 16 using atom-economical and chemoselective approaches

Barry M. Trost[1] & Guangbin Dong[1]

Of the concepts used to improve the efficiency of organic syntheses, two have been especially effective: atom economy[1] (the use of routes in which most of the atoms present in the reactants also end up in the product) and chemoselectivity[2] (the use of reactions that take place only at desired positions in a molecule). Synthesis of complex natural products is the most demanding arena in which to explore such principles. The bryostatin family of compounds are especially interesting targets, because they combine structural complexity with promising biological activity[3–7]. Furthermore, synthetic routes to some bryostatins have already been reported[9–12], providing a benchmark against which new syntheses can be measured. Here we report a concise total synthesis of bryostatin 16 (1), a parent structure from which almost all other bryostatins could in principle be accessed. Application of atom-economical and chemoselective reactions currently under development provides ready access to polyhydropyran motifs in the molecule, which are common structural features of many other natural products. The most notable transformations are two transition-metal-catalysed reactions. The first is a palladium-catalysed reaction of two different alkynes to form a large ring. The product of this step is then converted into a dihydropyran (the 'C ring' of bryostatins) in the second key reaction, which is catalysed by a gold compound. Analogues of bryostatin that do not exist in nature could be readily made by following this route, which might allow the biological activity of bryostatins to be fine-tuned.

The bryostatins 1–20 (Fig. 1), which were originally isolated from the marine bryozoan *Bugula neritina*, are a class of structurally complex macrolactone natural products that exhibit exceptional biological activity, most notable their anticancer activity *in vivo*[3]. Clinical application of bryostatin in combinatizon with other chemotherapeutic agents has shown significant potential in treating some cancers with high potency[4,5]. Furthermore, recent studies have revealed that bryostatin significantly affects both cognition and memory enhancement in animals, which suggests its potential use in the treatment of Alzheimer's disease, depression and other cognitive impairments[6]. Although bryostatins' activities could be attributed to their strong affinity for protein kinase C isozymes[7], their actual mode of action is still an important research subject. However, their clinical advancement is hampered by the limited availability of bryostatins from isolation, due to low yield ($\sim 1.6 \times 10^{-4}$%; 18 g of bryostatin 1, one of the most abundant members, from 14 t of animals on an industrial scale[8]) and there being a non-renewable supply. Therefore, efficient total syntheses of these natural products[9–12] and their analogues[13–16] remain in high demand.

The structures of bryostatins pose significant challenges to their synthesis; the structures include three heavily substituted tetrahydropyran rings, two acid/base-sensitive *exo*-cyclic unsaturated esters and one congested C16–C17 *trans*-alkene, as well as numerous oxygen-containing functionalities on a 26-membered lactone. As
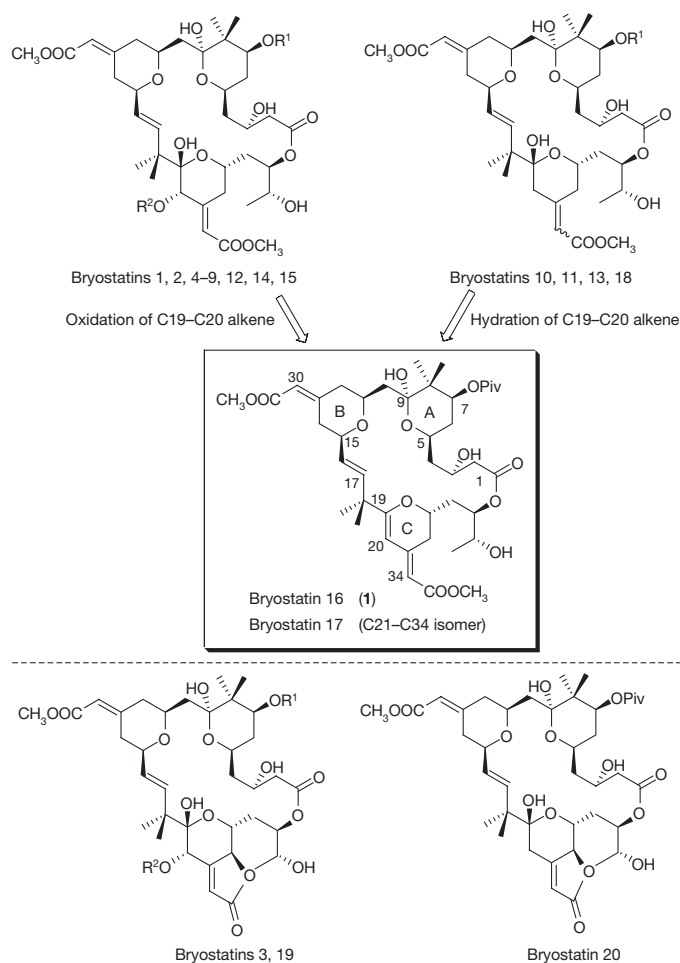
Figure 1 | Structures of bryostatins 1–20. Piv, pivaloyl.

an example, the challenge posed by the C16–C17 double bond led to failure in routes relying on its formation by metathesis reactions[17], even in the case of a relay metathesis strategy[16]. Despite their biological, clinical and structural significance, until now only three of the 20 bryostatins have been accessed by total synthesis (bryostatin 7 (ref. 9), bryostatin 2 (ref. 10) and bryostatin 3 (ref. 11)).

With a goal of streamlining the strategy to these complicated targets to enable better access, we choose bryostatin 16 (1)[18] as the specific synthetic target, for three reasons. First, bryostatin 16 could act as a pivotal parent structure allowing access to all other bryostatins (except bryostatins 3, 19 and 20; see Fig. 1) by elaboration of the electron-rich and relatively reactive C19–C20 alkene[10,16] (for a biosynthetic approach, see ref. 19). Second, the dihydropyran entity of

[1]Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA.

the C ring in **1** offers an ideal forum for us to explore a palladium-catalysed alkyne–alkyne coupling as a macrocyclization method[20] for complex natural product synthesis. Third, new analogues, which are not easily available from other syntheses, might be readily obtained simply by variations in this natural product's synthesis.

From the viewpoint of retrosynthetic analysis, the acid and/or base sensitivity of the C ring of bryostatins[10] leads us to a strategy of constructing the C ring of bryostatin 16 at the very end of the synthesis. The benefits of this also include flexible late variations for access to other bryostatins or analogues, as well as minimization of functional group transformation and protecting group usage. Whereas all previous total syntheses have relied on assembling the macrocycle by performing a difficult Julia olefination followed by a lactonization, we predict that the use of a palladium-catalysed chemoselective alkyne–alkyne coupling followed by a metal-catalysed 6-*endo-dig* cyclization will efficiently produce both the macrocycle and the C ring of **1** (Fig. 2). Esterification between fragments **4** and **5** will give the requisite diyne precursor. Fragment **5** can be synthesized from vinyl silane **6**. The 4-methylene-2,6-*cis*-tetrahydropyran moiety in intermediate **6** provides us with an opportunity to examine our ruthenium-catalysed alkene–alkyne coupling/Michael addition methodology[21] between two complex fragments (**7** and **8**), with the aim of addressing some questions in chemoselectivity. Given the difficulty of forming the sterically hindered C16–C17 alkene in the late stage (either by Julia olefination[10] or ring-closing metathesis[16,17]), alkyne **8** is specifically designed to install this *trans*-alkene in an early stage.

Alkene **7** has previously been synthesized in 16 steps from (*R*)-pantolactone[16]. This synthesis can be shortened (Fig. 3), however, by
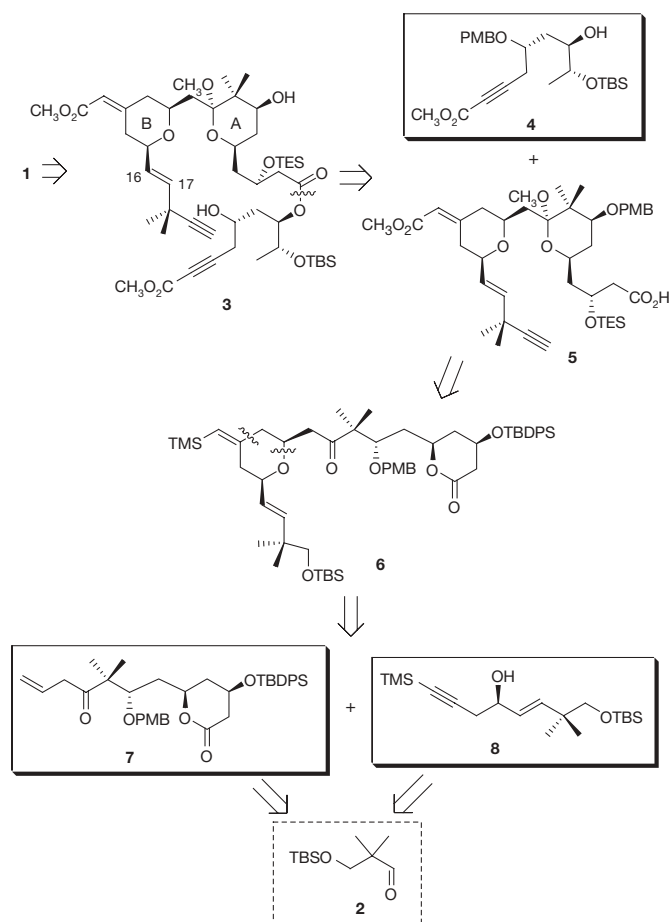
starting from aldehyde **2**. Asymmetric Brown allylation[22], followed by protection of the resulting alcohol with *p*-methoxybenzyl (PMB) group and oxidative alkene-cleavage, quickly afforded aldehyde **10** as the same intermediate in our previous synthesis. With this modification, alkene **7** is now available in 11 steps from aldehyde **2**. Enantioselective synthesis of alkyne **8** is achieved in four steps from aldehyde **2** (Fig. 3). Homologation, followed by indium-mediated propargylation[23] efficiently gives racemic **8** in good yield. (*R*)-**8** is then obtained in 90% yield and 90% enantiomeric excess through careful Dess–Martin oxidation[24] followed by Corey–Bakshi–Shibata reduction[25] of the corresponding ketone.

With both alkene **7** and alkyne **8** in hand, we proceed with the ruthenium-catalysed tandem alkene–alkyne coupling/Michael addition to generate *cis*-tetrahydropyran **6**. The chemoselectivity is demonstrated by the high compatibility of a β,γ-unsaturated ketone, a six-membered lactone, an unprotected allylic alcohol, a PMB ether and two different silyl ethers in this reaction. We find dichloromethane (DCM) to be the optimal solvent for this reaction, whereas acetone or a dichloromethane–*N,N*-dimethylformamide mixed solvent gives either lower conversion or more decomposition. Notably, only 1.2 equiv. of alkene **7** is required in this coupling reaction (Fig. 4a). Although the yield is moderate, presumably because additional alkene functionality in the alkyne fragment could lower the turnover number of the ruthenium catalyst, this result has been proved to be highly reproducible and both starting materials could be recovered, guaranteeing enough materials for the rest of the synthesis. Subsequent bromination of the *exo*-cyclic vinyl silane followed by a camphorsulfonic-acid-catalysed transesterification/methyl ketalization/desilylation all in one event cleanly gives the desired alcohol **13** containing both the A-ring and B-ring substructures in over 90% yield. The vinyl bromide functionality may serve as a convenient handle for the syntheses of bryostatin analogues through the use of metal-catalysed coupling reactions. We next use palladium-catalysed carbonylation to install the *exo*-cyclic conjugated methyl ester. Dess–Martin oxidation of the primary alcohol **14** followed by Ohira–Bestmann alkynylation[26] and desilylation provides donor alkyne **15** for the alkyne–alkyne coupling. We overcome the challenge of chemoselective hydrolysis of the β-hydroxy methyl ester in the presence of the α,β-unsaturated methyl ester by using of trimethyltin hydroxide[27] in 1,2-dichloroethane; we hypothesize that, owing to the Lewis acidity of trimethyltin hydroxide, the adjacent alcohol could act as a directing group in the saponification reaction. Subsequent protection of the alcohol with triethylsilyl group completes the synthesis of acid fragment **5**. Alcohol fragment **4** is synthesized in three steps from the known[16] homo-propargyl alcohol **17** (Fig. 4b).
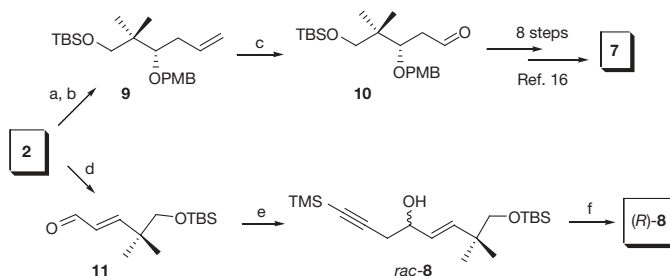
Figure 3 | Synthesis of alkene 7 and alkyne 8. Reaction conditions: (a) (−)-(Ipc)$_2$B(allyl), Et$_2$O, −90 °C, 67%, 94% enantiomeric excess (e.e.); (b) PMB-Br, NaH, DMF, 90%; (c) OsO$_4$ (2 mol%), 2,6-lutidine, NaIO$_4$, dioxane/water (3:1), 87%; (d) (*Z*)-1-bromo-2-ethoxyethene, *t*-butyllithium, (CH$_3$)$_2$Zn, then **2**, (C$_2$H$_5$)$_2$O, −78 °C; then NaHSO$_4$, room temperature (20–25 °C), 97%; (e) (3-bromo-1-propynyl)-trimethylsilane, indium powder, InF$_3$ (10 mol%), THF, 65 °C, 68%; (f)(i) Dess–Martin periodinane, NaHCO$_3$, DCM; (f)(ii) (*S*)-2-methyl-CBS-oxazaborolidine (5 mol%), catecholborane, DCM, −78 °C, 90%, 90% e.e. over two steps. Ipc, isopinocamphenyl; DMF, *N,N*-dimethylformamide; THF, tetrahydrofuran; DCM, dichloromethane; CBS, Corey–Bakshi–Shibata. For tabulated spectral data of all depicted compounds, please see the Supplementary Information.

Figure 2 | Retrosynthetic analysis. TES, triethylsilyl; TBS, *t*-butyldimethylsilyl; PMB, *p*-methoxybenzyl; TMS, trimethylsilyl; TBDPS, *tert*-butyldiphenylsilyl.

Esterification between acid **5** and alcohol **4** proceeds in 92% yield using Yamaguchi's conditions[28] (Fig. 5). Oxidative removal of the two PMB protecting groups using 2,3-dichloro-5,6-dicyano-1,4-benzoquinone gives the macrocyclization precursor **3**. After extensive experimentation, we find that treatment of **3** with 12 mol% $Pd(OAc)_2$ and 15 mol% tris(2,6-dimethoxyphenyl)phosphine in toluene at room temperature (20–25 °C) successfully provides the desired macrocycle **20** with reasonably good yield (56%), whereas the use of tetrahydrofuran or benzene as solvent, or a lower ligand/palladium ratio proves less efficient. As for other macrocyclizations, low concentration (0.002 M) proves to be critical; otherwise, formation of the dimeric by-products could be observed. This example of using palladium-catalysed alkyne–alkyne coupling as a macrocyclization method in a complex natural product synthesis illustrates a new way of constructing a macrocycle using carbon–carbon bond formation. Mechanistically, the palladium catalyst chemoselectively inserts into the carbon–hydrogen bond of the terminal alkyne; this sets the stage for the chemo- and regioselective intramolecular

carbometalation of the disubstituted alkyne, which, after reductive elimination of the formed vinyl palladium hydride, creates the macrocycle efficiently in spite of the complexity of the substrate.

The remaining challenge is to conduct a 6-*endo-dig* cyclization to form the C ring of bryostatin. Owing to the modest selectivity in the palladium-catalysed reaction (5-*exo* versus 6-*endo*) and the difficulty in separating these isomers[16], we seek a more regioselective catalyst. After extensive screening of a number of metals, we settle upon a cationic gold complex, $[Au(PPh_3)]SbF_6$, as the catalyst (for a cationic gold-catalysed 5-*exo* cyclization, see ref. 29). DCM-CH_3CN (10:1), as a mixed solvent in the presence of $NaHCO_3$ (10 equiv.) as a buffer, gives the acid-sensitive 6-*endo* product in 73% isolated yield. Subsequent pivalation of the hindered secondary alcohol under forcing conditions (Piv_2O, 50 equiv.; *N*,*N*-4-dimethylaminopyridine, 80 equiv.; 50 °C)[30] does afford the pivalate ester **21** in 62% yield. The following global deprotection proves to be nontrivial: hydrogen fluoride-pyridine, aqueous hydrofluoric acid, aqueous acetic acid, pyridinium *p*-toluenesulfonate and so on give either decomposition or isomerization. We eventually find an extreme acid sensitivity of this natural product. By contrast with these acid conditions, treatment of **21** with 5 equiv. of tetra-*n*-butylammonium fluoride and direct purification by reverse-phase high-performance liquid chromatography successfully provides bryostatin 16 (**1**) that is spectroscopically identical to previously reported samples (reported optical
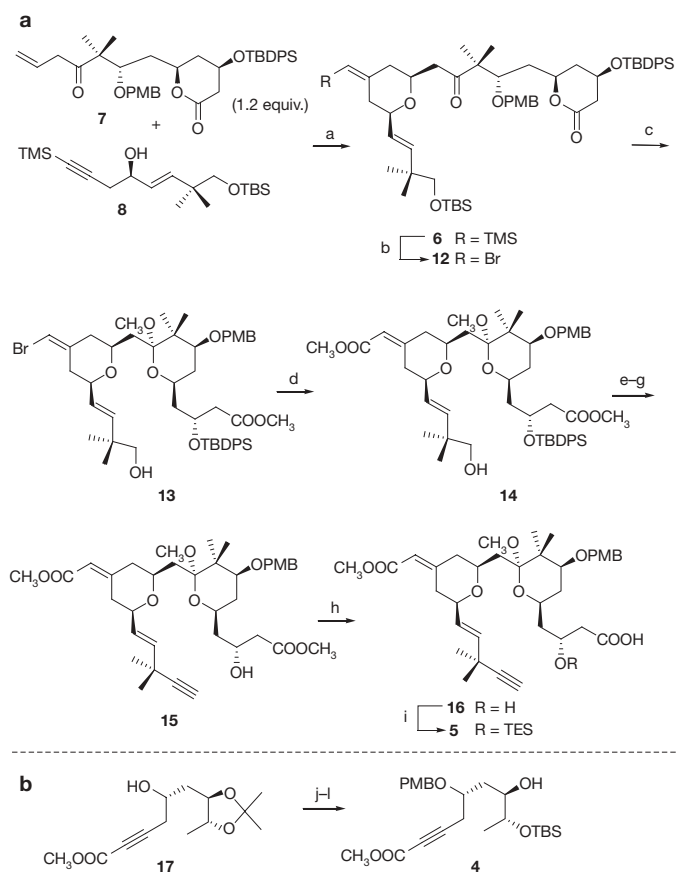
Figure 4 | Synthesis of acid 5 and alcohol 4. a, Synthesis of **5**. Reaction conditions: (a) $CpRu(CH_3CN)_3PF_6$ (13 mol%), DCM, 34% (80% b.r.s.m.); (b) NBS, DMF, 98%; (c) CSA (10 mol%), CH_3OH, 0 °C, 93–96%; (d) $PdCl_2(CH_3CN)_2$ (10 mol%), dppf (30 mol%), CO (1 atm), CH_3OH, $(C_2H_5)_3N$, DMF, 80 °C, 83% (90% b.r.s.m.); (e) Dess–Martin periodinane, $NaHCO_3$, DCM, 88%; (f) Ohira–Bestmann reagent, K_2CO_3, CH_3OH, 97%; (g) TBAF, HOAc, THF, 90% (96% b.r.s.m.); (h) $(CH_3)_3SnOH$, DCE, 80 °C, 84%; (i) TESOTf, 2,6-lutidine, DCM, −10 °C to 0 °C, 76–79%. **b**, Synthesis of **4**. Reaction conditions: (j) $Cu(OTf)_2$ (3 mol%), PMBOC(NH)CCl_3, toluene, −10 °C; (k) PPTS, CH_3OH, 93% over two steps; (l) TBSOTf, 2,6-lutidine, DCM, −78 °C, 71%. Cp, cyclopentadienyl; b.r.s.m., based on recovered starting material; NBS, *N*-bromosuccinimide; CSA, camphorsulfonic acid; dppf, 1,1′-bis(diphenylphosphino)ferrocene; TBAF, tetra-*n*-butylammonium fluoride; HOAc, acetic acid; DCE, 1,2-dichloroethane; OTf, trifluoromethanesulfonate; PPTS, pyridinium *p*-toluenesulfonate. For tabulated spectral data of all depicted compounds, please see the Supplementary Information.
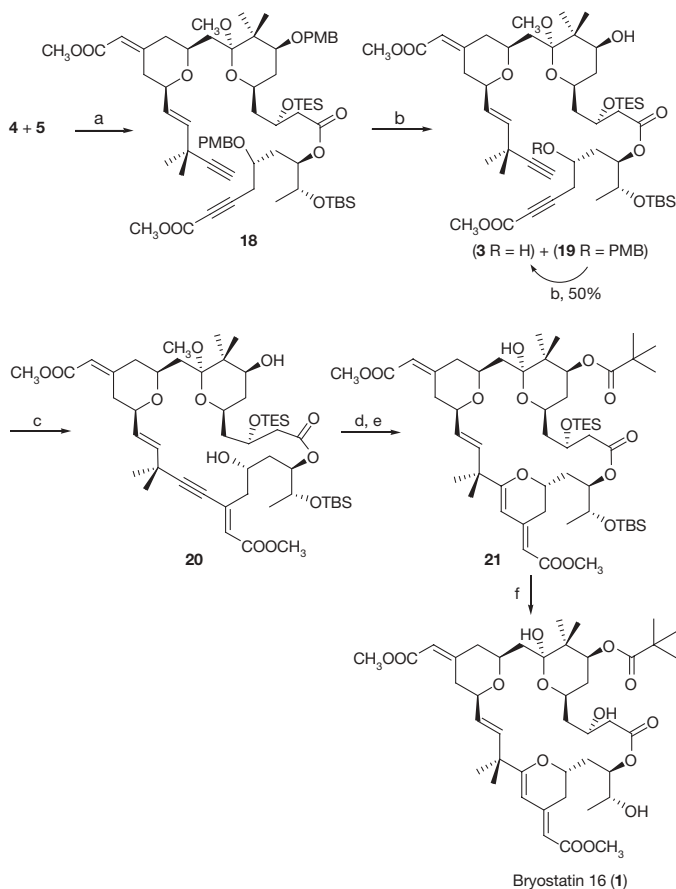
Figure 5 | Synthesis of bryostatin 16. Reaction conditions: (a) **5**, 2,4,6-trichlorobenzoyl chloride, $(C_2H_5)_3N$, toluene, then **4**, DMAP, 92%; (b) DDQ, pH 7.0 buffer, DCM, 46% **3** and 58% **19**; (c) $Pd(OAc)_2$ (12 mol%), TDMPP (15 mol%), toluene, 56%; (d) $AuCl(PPh_3)$ (20 mol%), $AgSbF_6$ (20 mol%), $NaHCO_3$, DCM/CH_3CN, 0 °C to room temperature, 73%; (e) Piv_2O, DMAP, DCM, 50 °C, 62%; (f) TBAF, THF, ~52%. DMAP, *N*,*N*-4-dimethylaminopyridine; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; TDMPP, tris(2,6-dimethoxyphenyl)phosphine. For tabulated spectral data of all depicted compounds, please see the Supplementary Information.

487

rotation: $[\alpha]_D^{20}$, $+84°\,cm^3\,g^{-1}\,dm^{-1}$ ($c = 0.43\,g\,cm^{-3}$, in CH_3OH); found: $[\alpha]_D^{20}$, $+81°\,cm^3\,g^{-1}\,dm^{-1}$ ($c = 0.04\,g\,cm^{-3}$, in CH_3OH)[18].

In conclusion, we have developed a highly concise strategy (26-step longest linear sequence, 39 total steps from aldehyde **2**) for the asymmetric total synthesis of bryostatin 16. The synthetic efficiency can be attributed to a tandem ruthenium-catalysed alkene–alkyne coupling/Michael addition to form the B ring, an acid-catalysed one-pot cascade to form the A ring, a directed chemoselective hydrolysis, a palladium-catalysed alkyne–alkyne coupling as a macrocyclization reaction and a gold-catalysed 6-*endo-dig* cyclization to form the C ring of bryostatin 16. We believe that all these atom-economical and chemoselective approaches could have implications beyond this work. The conciseness of this synthesis readily allows access to significant quantities of this key bryostatin, and implementation of this strategy towards the synthesis of various bryostatins and their analogues, and performance of the related biological experiments, is being undertaken.

1. Trost, B. M. The atom economy–a search for synthetic efficiency. *Science* **254**, 1471–1477 (1991).
2. Trost, B. M. Selectivity: a key to synthetic efficiency. *Science* **219**, 245–250 (1983).
3. Hale, K. J., Hummersone, M. G., Manaviazar, S. & Frigerio, M. The chemistry and biology of the bryostatin antitumour macrolides. *Nat. Prod. Rep.* **19**, 413–453 (2002).
4. Newman, D. J. & Cragg, G. M. Marine natural products and related compounds in clinical and advanced preclinical trials. *J. Nat. Prod.* **67**, 1216–1238 (2004).
5. Dowlati, A. *et al.* Phase I and correlative study of combination bryostatin 1 and vincristine in relapsed B-cell malignancies. *Clin. Cancer Res.* **9**, 5929–5935 (2003).
6. Hongpaisan, J. & Alkon, D. L. A structural basis for enhancement of long-term associative memory in single dendritic spines regulated by PKC. *Proc. Natl Acad. Sci. USA* **104**, 19571–19576 (2007).
7. Dell'Aquila, M. L., Harold, C. L., Kamano, Y., Pettit, G. R. & Blumberg, P. M. Differential effects of bryostatins and phorbol esters on arachidonic acid metabolite release and epidermal growth factor binding in C3H 10T1/2 cells. *Cancer Res.* **48**, 3702–3708 (1988).
8. Schaufelberger, D. E. *et al.* The large-scale isolation of bryostatin 1 from Bugula neritina following current good manufacturing practices. *J. Nat. Prod.* **54**, 1265–1270 (1991).
9. Kageyama, M. *et al.* Synthesis of bryostatin 7. *J. Am. Chem. Soc.* **112**, 7407–7408 (1990).
10. Evans, D. A. *et al.* Total synthesis of bryostatin 2. *J. Am. Chem. Soc.* **121**, 7540–7552 (1999).
11. Ohmori, K. *et al.* Total synthesis of bryostatin 3. *Angew. Chem. Int. Ed.* **39**, 2290–2294 (2000).
12. Manaviazar, S. *et al.* Enantioselective formal total synthesis of the antitumor macrolide bryostatin 7. *Org. Lett.* **8**, 4477–4480 (2006).
13. Wender, P. A. & Verma, V. A. The design, synthesis and evaluation of C7 diversified bryostatin analogs reveals a hot spot for PKC affinity. *Org. Lett.* **10**, 3331–3334 (2008).
14. Wender, P. A., DeChristopher, B. A. & Schrier, A. J. Efficient synthetic access to a new family of highly potent bryostatin analogues via a Prins-driven macrocyclization strategy. *J. Am. Chem. Soc.* **130**, 6658–6659 (2008).
15. Keck, G. E. *et al.* Convergent assembly of highly potent analogues of bryostatin 1 via pyran annulation: bryostatin look-alikes that mimic phorbol ester function. *J. Am. Chem. Soc.* **130**, 6660–6661 (2008).
16. Trost, B. M., Yang, H., Thiel, O. R., Frontier, A. J. & Brindle, C. S. Synthesis of a ring-expanded bryostatin analogue. *J. Am. Chem. Soc.* **129**, 2206–2207 (2007).
17. Ball, M. *et al.* A preliminary evaluation of a metathesis approach to bryostatins. *Tetrahedr. Lett.* **47**, 2223–2227 (2006).
18. Pettit, G. R. *et al.* Antineoplastic agents. 340. Isolation and structural elucidation of bryostatins 16–18. *J. Nat. Prod.* **59**, 286–289 (1996).
19. Sudek, S. *et al.* Identification of the putative bryostatin polyketide synthase gene cluster from "Candidatus Endobugula sertula", the uncultivated microbial symbiont of the marine ryozoan Bugula neritina. *J. Nat. Prod.* **70**, 67–74 (2007).
20. Trost, B. M., Matsubara, S. & Caringi, J. J. Cycloisomerization of α,ω-diynes to macrocycles. *J. Am. Chem. Soc.* **111**, 8745–8746 (1989).
21. Trost, B. M., Yang, H. & Wuitschik, G. A. Ru-catalyzed tandem alkyne-enone coupling/Michael addition: synthesis of 4-methylene-2,6-*cis*-tetrahydropyrans. *Org. Lett.* **7**, 4761–4764 (2005).
22. Brown, H. C. & Jadhav, P. K. Asymmetric carbon-carbon bond formation via β-allyldiisopinocampheylborane. Simple synthesis of secondary homoallylic alcohols with excellent enantiomeric purities. *J. Am. Chem. Soc.* **105**, 2092–2093 (1983).
23. Lin, M. & Loh, T. Indium-mediated reaction of trialkylsilyl propargyl bromide with aldehydes: highly regioselective synthesis of allenic and homopropargylic alcohols. *J. Am. Chem. Soc.* **125**, 13042–13043 (2003).
24. Dess, D. B. & Martin, J. C. A useful 12-I-5 triacetoxyperiodinane (the Dess-Martin periodinane) for the selective oxidation of primary or secondary alcohols and a variety of related 12-I-5 species. *J. Am. Chem. Soc.* **113**, 7277–7287 (1991).
25. Corey, E. J. & Helal, J. C. Reduction of carbonyl compounds with chiral oxazaborolidine catalysts: A new paradigm for enantioselective catalysis and a powerful new synthetic method. *Angew. Chem. Int. Ed.* **37**, 1986–2012 (1998).
26. Roth, G. J., Liepold, B., Müller, S. G. & Bestmann, H. J. Further improvements for the synthesis of alkynes from aldehydes. *Synthesis* 59–62 (2004).
27. Nicolaou, K. C., Estrada, A. A., Zak, M., Lee, S. H. & Safina, B. S. A mild and selective method for the hydrolysis of esters with trimethyltin hydroxide. *Angew. Chem. Int. Ed.* **44**, 1378–1382 (2005).
28. Inanaga, J., Hirata, K., Saeki, H., Katsuki, T. & Yamaguchi, M. A rapid esterification by mixed anhydride and its application to large-ring lactonization. *Bull. Chem. Soc. Jpn* **52**, 1989–1993 (1979).
29. Liu, Y., Song, F., Song, Z., Liu, M. & Yan, B. Gold-catalyzed cyclization of (Z)-2-en-4-yn-1-ols: highly efficient synthesis of fully substituted dihydrofurans and furans. *Org. Lett.* **7**, 5409–5412 (2005).
30. Pettit, G. R., Sengupta, D., Herald, C. L., Sharkey, N. A. & Blumberg, P. M. Antineoplastic agents. 192. Synthetic conversion of bryostatin 2 to bryostatin 1 and related bryopyrans. *Can. J. Chem.* **69**, 856–860 (1991).

Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation

A. Biastoch[1], C. W. Böning[1] & J. R. E. Lutjeharms[2]

Predicting the evolution of climate over decadal timescales requires a quantitative understanding of the dynamics that govern the meridional overturning circulation (MOC)[1]. Comprehensive ocean measurement programmes aiming to monitor MOC variations have been established in the subtropical North Atlantic[2,3] (RAPID, at latitude 26.5° N, and MOVE, at latitude 16° N) and show strong variability on intraseasonal to interannual timescales. Observational evidence of longer-term changes in MOC transport remains scarce, owing to infrequent sampling of transoceanic sections over past decades[4,5]. Inferences based on long-term sea surface temperature records, however, supported by model simulations, suggest a variability with an amplitude of ±1.5–3 Sv ($1 \, \mathrm{Sv} = 10^6 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$) on decadal timescales in the subtropics[6]. Such variability has been attributed to variations of deep water formation in the sub-arctic Atlantic, particularly the renewal rate of Labrador Sea Water[7]. Here we present results from a model simulation that suggest an additional influence on decadal MOC variability having a Southern Hemisphere origin: dynamic signals originating in the Agulhas leakage region at the southern tip of Africa. These contribute a MOC signal in the tropical and subtropical North Atlantic that is of the same order of magnitude as the northern source. A complete rationalization of observed MOC changes therefore also requires consideration of signals arriving from the south.

The Agulhas leakage around South Africa[8] is the main source of the warm and salty waters carried towards the subpolar North Atlantic as the upper limb of the MOC. The Agulhas leakage region is characterized by vigorous variability on intraseasonal to interannual timescales[9] shedding the largest mesoscale eddies in the world ocean[10], which are the dominating vehicles transporting and releasing the Indian Ocean waters into the Atlantic[11]. Observational palaeoclimate studies have suggested a broken inter-ocean exchange on centurial to millennial timescales[12], and ocean climate model studies[13,14] have elucidated the potential of the inter-ocean exchange of heat and salt to alter the long-term MOC response. However, large uncertainties in those studies remain, owing to unresolved mesoscale processes[15] that govern the dynamics and strength of the Agulhas leakage[16]; the present study focuses on those effects on decadal timescales.

The methodology used here is to assess the effect of mesoscale processes in the Agulhas leakage by adopting a two-way nesting scheme that selectively increases the grid resolution in the region of interest in a global model simulation. More specifically, we use the following approach. We start with a global ocean/sea-ice model with a nominal grid resolution of 0.5° (ORCA05) that reasonably resolves western boundary current structures and captures the different processes of deep-water formation due to the thermohaline and wind forcing, but that does not resolve mesoscale processes. ORCA05 has been demonstrated[17] to capture the decadal MOC variability (Fig. 1) simulated in more highly resolved, eddying models; it thus serves as a

meaningful basis for the present study. In a second configuration (AG01), we selectively refine the resolution (to 0.1°) of the Agulhas region in this model by adopting a 'two-way nesting' scheme (Fig. 2a). The high-resolution nest provides a realistic simulation of the mesoscale dynamics of the Agulhas retroflection (Fig. 2b, c), including the upstream perturbations originating in the source regions of the Agulhas current[15] that control the frequency of the inter-ocean exchange[16]. The effect of these processes is fed back to the global model[16,18], allowing the global circulation to respond to the mesoscale dynamics introduced by this limited region. Comparison of the two model simulations, one with an Agulhas nest (AG01) and one without (ORCA05), thus provides us with an effective means of identifying the impact of the mesoscale Agulhas dynamics on the basin-scale MOC. The alternative, that is, analysis of a globally eddying model[19], would provide a comprehensive hindcast of MOC variability, but would not allow us to discriminate effects due to the mesoscale processes of an individual region (the Agulhas leakage region in this case) from those arising in other areas. In addition, local variability[17] would further complicate the interpretation.

Although the introduction of the high-resolution Agulhas nest has little effect on the time-mean MOC transport (differences are less than 1 Sv; Supplementary Fig. 2), despite the large difference in the mean inter-oceanic transport from the Indian Ocean to the Atlantic with realistic[11] transport of 12 Sv in AG01 and an over-estimation of 21.5 Sv in ORCA05[16], there is a marked difference in the MOC variability (Fig. 3): an isolation of the MOC variability induced by the mesoscale fluctuations in the Agulhas (that is, the 'Agulhas-induced MOC anomalies' found by comparing the experiments with and
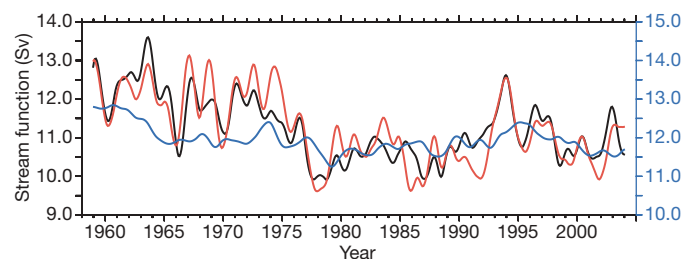

Figure 1 | Strength of the interhemispheric transport in the Atlantic Ocean. Shown is the low-pass-filtered (23-month Hanning filter) maximum stream function value of the North Atlantic Deep Water cell in the global coarse-resolution model without (ORCA05, black) and with (AG01, red) a high-resolution Agulhas nest (left-hand scale). The blue curve (right-hand scale (Sv)) shows a sensitivity experiment with the global coarse-resolution model in which only heat fluxes varied interannually (HEAT[17]). For wind stress and freshwater fluxes, a repeated year was used (that is, still with daily forcing, but without interannual variation). We note the slightly different strength of the stream function in the sensitivity experiment, which is due to a general off-set between the climatological and interannually varying forcing[27].

[1]Leibniz-Institut für Meereswissenschaften, Düsternbrooker Weg 20, 24105 Kiel, Germany. [2]Department of Oceanography, University of Cape Town, 7700 Rondebosch, South Africa.
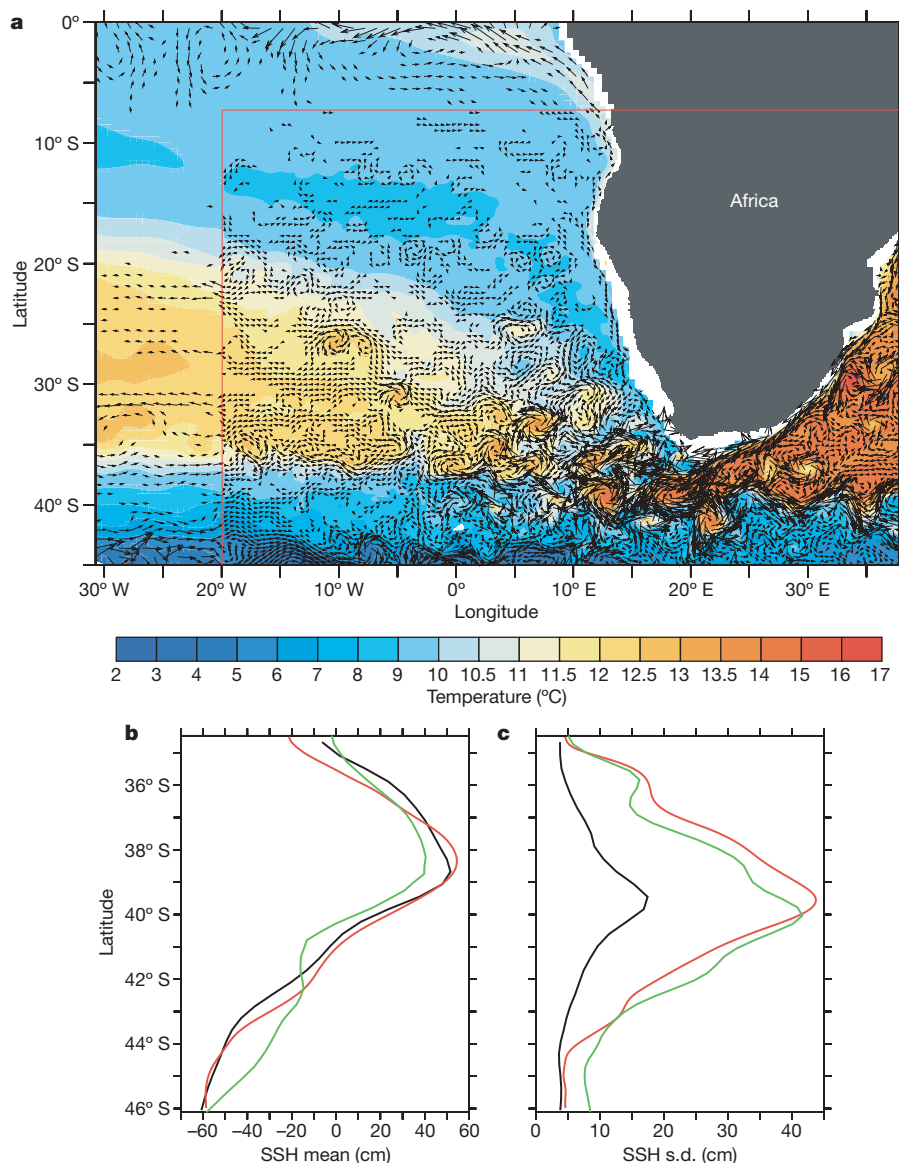
Figure 2 | Mid-depth circulation in the high-resolution Agulhas nest.
a, Temperature (colour) and velocities (every fourth vector shown of a five-day average around 15 September 2004) at 450-m depth for model experiment AG01, the high-resolution (0.1°) Agulhas nest (the northwestern portion is delineated by the red lines) hosted in a global ocean model of 0.5° resolution. **b**, **c**, Mean (**b**) and standard deviation (**c**) of sea surface height (SSH) at longitude 20° E of the global coarse-resolution model (ORCA05, black), the high-resolution Agulhas nest (AG01, red) and satellite data (Aviso, green).

without the Agulhas nest) illustrates an effect on interannual to decadal timescales. There is a remarkable meridional coherence in the Agulhas-induced MOC signal: anomalies of more than ±1.5 Sv emerge near latitude 30° S and rapidly propagate towards the equator, with only slight damping. In the Northern Hemisphere the anomalies then gradually fade; at 20° N, amplitudes are below 0.5 Sv. The propagation from 30° S into the Northern Hemisphere is very rapid, occurring on the order of months, which is indicative of Kelvin wave processes as discussed in previous ocean model studies[20,21] and climate models[22]. The propagation is thus about 1–2 orders of magnitude faster than it would be if it were due to advective processes, such as the translation of Agulhas rings (which occurs at a few centimetres per second)[23]. This solution suggests that wave processes along the western boundary of the Atlantic Ocean have a prime role in the rapid communication of the signal, a mechanism (Fig. 4) that was theoretically described in idealized model studies[24].

Agulhas rings are the dominating vehicle of the inter-basin exchange into the Cape Basin off the west coast of South Africa. They can be reasonably traced by the depth of the 10 °C isotherm[25],

where the anticyclonic rotation sense of a ring causes depressions in the isothermal surfaces. Time series of isopycnal depth along 30° S at full temporal resolution (Supplementary Fig. 4) indicate westward propagating signals of about 6 cm s^{-1}, in good agreement with observations of Agulhas rings[23]. The signals reveal a decadal modulation (Supplementary Fig. 5) that is also evident in the sea surface height, in good agreement with the observational record (Supplementary Fig. 6). The modulation has a similar temporal characteristic as the Agulhas-induced MOC signal, suggesting a projection of the zonally propagating, meridional velocity anomalies onto the zonally integrated MOC transport in this latitude range. Farther north, the MOC variability signal is mainly concentrated at the western boundary, which can be seen in a comparison of the MOC transport anomalies with the transport anomalies of the North Brazil current (Fig. 3a).

How significant are the Agulhas-induced MOC anomalies and what is their relative importance in comparison with decadal MOC signals originating from the subpolar North Atlantic? Although changes in the dense overflow at the Greenland–Scotland ridge
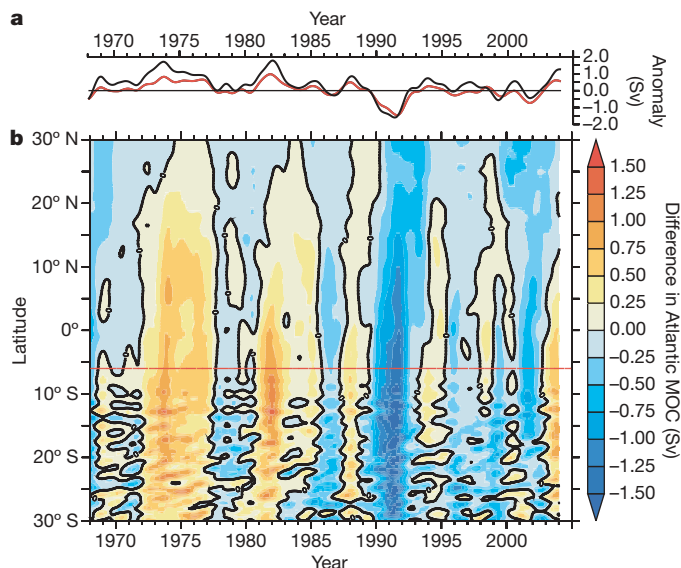
Figure 3 | Low-pass-filtered Agulhas-induced MOC anomalies.
a, Comparison of the (low-pass-filtered) time series of the MOC anomalies (red) at 6° S (dashed red line in **b**) with the corresponding anomalies of the North Brazil current at this latitude (black). **b**, Hovmoeller diagram showing the difference in Atlantic MOC between model experiments with (AG01) and without (ORCA05) a high-resolution Agulhas nest. Integrated over the upper 1,000 m, this typically represents the total strength of the MOC (compare with Supplementary Fig. 2).

appear to have had comparatively minor effects in past decades[6], modifications in the formation of Labrador Sea Water were identified in model studies as an important source of basin-scale MOC signals on decadal timescales[7], leading to mid-latitude North Atlantic MOC changes on the order of 1–2 Sv.

A quantitative assessment of the relative importance of the various MOC signals is provided in Fig. 5. In the experiments with (red) and without (black) the Agulhas nest, the total variability is broadly similar, owing to the dominance of local wind-induced variability on interannual timescales[17]. The individual effect of the mesoscale Agulhas dynamics (red), given by the difference of the two MOC time series, has a maximum amplitude in the South Atlantic, but reaches well into the tropical (North) Atlantic. For comparison, the figure also depicts the influence due to the variability in sub-arctic deep-water formation: in a previous study[17] its effect was isolated by a sensitivity experiment
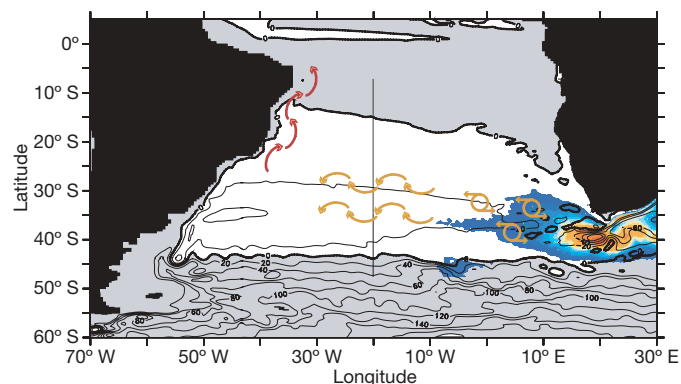
Figure 4 | Illustration of the wave processes conveying Agulhas-induced anomalies in the upper limb of the MOC. The contour lines depict the time-mean barotropic stream function (Sv), indicating the anticyclonic (white area) subtropical gyre in the South Atlantic; the colour information shows the time-mean eddy kinetic energy. The arrows give an illustration of the dynamic processes transporting eddies and Rossby waves originating in the Agulhas across the South Atlantic (yellow) and Kelvin waves along the continental slope of South America (red).
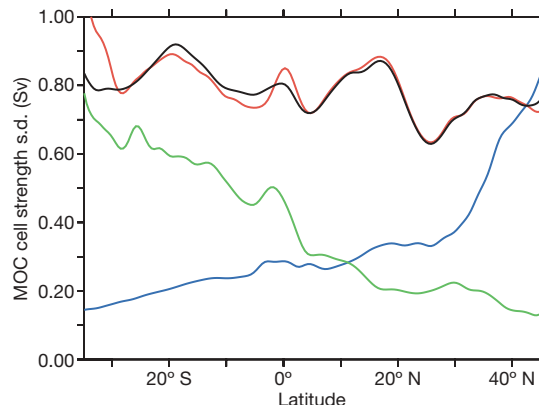
Figure 5 | Attribution of interannual MOC variability to different mechanisms. Standard deviation of the low-pass-filtered deep-water cell strength as a function of latitude, in the experiments without (ORCA05, black) and with (AG01, red) an Agulhas nest. The effect of the Agulhas-induced variability is isolated by the standard deviation of the MOC difference the two experiments (green). For comparison, the blue curve shows the individual effect of sub-arctic deep-water formation variability, obtained in a sensitivity experiment (HEAT[17]) in which ORCA05 was selectively driven by interannually varying heat-flux forcing and repeated-year wind and freshwater forcing.

(HEAT) based on ORCA05 in which interannual forcing variations were retained in the heat fluxes and a repeated year was used for wind and freshwater fluxes. In that case, MOC variability is primarily induced in the sub-arctic North Atlantic; its strength decreases towards the Equator, becoming comparable to that of the Agulhas effect in the tropical Atlantic. The simulations thus suggest that the variability of inter-hemispheric transport (Fig. 1) arises from a combination of effects from the north as well as the south. It also suggests that both influencing factors need to be invoked to interpret MOC variations in the subtropical North Atlantic, as currently studied in multi-year measurement programs at 16° N (MOVE)[3] and 26° N (RAPID)[2,3].

Using a two-way nesting scheme that selectively resolves mesoscale dynamics in the greater Agulhas region, the model simulations demonstrate that the Agulhas leakage acts as the source of decadal MOC variability: low-frequency undulations in thermocline depth induced in the leakage regime project onto a dynamical signal carried across the South Atlantic by Rossby waves and into the North Atlantic by wave processes along the American continental slope. This is reflected in a decadal MOC transport signal that gradually diminishes from south to north, but has an amplitude in the tropical Atlantic of comparable magnitude to the effect of sub-arctic deep-water formation processes discussed in previous studies. The effect of the Agulhas leakage dynamics on decadal timescales identified here adds to previous palaeoclimate studies[13] concluding that changes in inter-basin heat and fresh water fluxes between the Indian and Atlantic Ocean can strongly influence the MOC on centurial and longer timescales; it demonstrates the need to realistically capture the mesoscale dynamics of this regime in projections of climate change over the Atlantic Ocean.

METHODS SUMMARY

The model is based on the NEMO ocean/sea-ice model (v2.3)[26]. The global base model (ORCA05) has a nominal resolution of 0.5°, and therefore does not resolve the oceanic mesoscale. Its state-of-the-art set-up (including partial bottom cells and advanced advection schemes) has been used by the DRAKKAR consortium in a wide range of applications. This base model hosts a high-resolution, 0.1°-resolution regional model of the Agulhas region (20° W–70° E, 47° S–7° S, AG01[16]; Fig. 2). Using an adaptive mesh refinement[18], it not only receives its boundary conditions from the base model but is also able to update it. This 'two-way nesting' approach makes it possible to study the effect of the high-resolution dynamics in the nest on the global circulation, the main element of this study. The model system is driven by a consistent data set[27] of six-hourly to daily, interannually varying wind and thermohaline surface forcing fields over

the period 1958–2004. For comparison we use an ORCA05-based sensitivity experiment of a previous study[17] (HEAT), in which only heat fluxes were allowed to vary interannually; wind stress and freshwater fluxes had no variations from year to year. In this case, MOC variations are reverberations of deep-water formation events in the subpolar North Atlantic. (We note that the sensitivity experiment differs slightly in its model set-up, in that it mainly uses an eddy parameterization[28] that is necessary for a realistic representation of the deep-water formation processes in the Southern Ocean.) Regarding the wave processes communicating the dynamical signal northward, we point out that NEMO is formulated on a C grid, on which Kelvin wave propagation is unaffected by horizontal resolution[29].

The model fields are compared with an altimeter product by Aviso. In this data set, GRACE (Gravity Recovery And Climate Experiment) satellite sea surface heights were subtracted, and a new mean dynamics topography (Rio05)[30] using *in situ* measurements from drifters and hydrography was then introduced. (The altimeter products were produced by Ssalto/Duacs and distributed by Aviso with support from CNES. Rio05 was produced by CLS Space Oceanography Division.)

1. Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L. & Roeckner, E. Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* **453**, 84–88 (2008).
2. Cunningham, S. A. *et al.* Temporal variability of the Atlantic meridional overturning circulation at 26.5° N. *Science* **317**, 935–938 (2007).
3. Kanzow, T., Send, U., Zenk, W., Chave, A. D. & Rhein, M. Monitoring the integrated deep meridional flow in the tropical North Atlantic: Long-term performance of a geostrophic array. *Deep-Sea Res. I* **53**, 528–546 (2006).
4. Bryden, H., Longworth, H. R. & Cunningham, S. A. Slowing of the Atlantic meridional overturning circulation at 25° N. *Nature* **438**, 655–657 (2005).
5. Wunsch, C. Mass and volume transport variability in an eddy-filled ocean. *Nature Geosci.* **1**, 165–168 (2008).
6. Latif, M. *et al.* Is the thermohaline circulation changing? *J. Clim.* **19**, 4631–4637 (2006).
7. Böning, C. W., Scheinert, M., Dengg, J., Biastoch, A. & Funk, A. Decadal variability of subpolar gyre transport and its reverberation in the North Atlantic overturning. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL026906 (2005).
8. Gordon, A. L. Inter-ocean exchange of thermocline water. *J. Geophys. Res.* **91**, 5037–5046 (1986).
9. Lutjeharms, J. R. E. *The Agulhas Current* (Springer, 2006).
10. Olson, D. B. & Evans, R. H. Rings of the Agulhas Current. *Deep-Sea Res. A* **33**, 27–42 (1986).
11. De Ruijter, W. P. M. *et al.* Dynamics, estimation and impact of South Atlantic inter-ocean exchange. *J. Geophys. Res.* **104**, 20885–20910 (1999).
12. Peeters, F. J. C. *et al.* Vigorous exchange between Indian and Atlantic Ocean at the end of the last five glacial periods. *Nature* **430**, 661–665 (2004).
13. Weijer, W., de Ruijter, W. P. M., Dijkstra, H. A. & van Leeuwen, P. J. Impact of interbasin exchange on the Atlantic overturning circulation. *J. Phys. Oceanogr.* **29**, 2266–2284 (1999).
14. Marsh, R., Hazeleger, W., Yool, A. & Rohling, E. J. Stability of the thermohaline circulation under millennial CO_2 forcing and two alternative controls on Atlantic salinity. *Geophys. Res. Lett.* **34**, doi:10.1029/2007GL027815 (2007).
15. Van Leeuwen, P. J., de Ruijter, W. P. M. & Lutjeharms, J. R. E. Natal pulses and the formation of Agulhas rings. *J. Geophys. Res.* **105**, 6425–6436 (2000).
16. Biastoch, A., Lutjeharms, J. R. E., Böning, C. W. & Scheinert, M. Mesoscale perturbations control inter-ocean exchange south of Africa. *Geophys. Res. Lett.* doi:10.1029/2008GL035132 (in the press).
17. Biastoch, A., Böning, C., Getzlaff, J., Molines, J.-M. & Madec, G. Causes of interannual - decadal variability in the meridional overturning circulation of the mid-latitude North Atlantic Ocean. *J. Clim.*. doi:10.1175/2008JCLI2404.1 (in the press) (2008).
18. Debreu, L., Vouland, C. & Blayo, E. AGRIF: Adaptive grid refinement in Fortran. *Computers Geosci.* **34**, 8–13 (2008).
19. Maltrud, M. E. & McClean, J. An eddy resolving global 1/10° ocean simulation. *Ocean Model.* **8**, 31–54 (2005).
20. Johnson, H. L. & Marshall, D. P. A theory for the surface Atlantic response to thermohaline variability. *J. Phys. Oceanogr.* **32**, 1121–1132 (2002).
21. Getzlaff, J., Böning, C. W., Eden, C. & Biastoch, A. Signal propagation related to the North Atlantic overturning. *Geophys. Res. Lett.* **32**, doi:10.1029/2004GL021002 (2005).
22. Dong, B. W. & Sutton, R. T. Adjustment of the coupled ocean-atmosphere system to a sudden change in the thermohaline circulation. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL015229 (2002).
23. Garzoli, S. L. *et al.* Three Agulhas rings observed during the Benguela Current Experiment. *J. Geophys. Res.* **104**, 20971–20985 (1999).
24. Van Sebille, E. & van Leeuwen, P. J. Fast northward energy transfer in the Atlantic due to Agulhas rings. *J. Phys. Oceanogr.* **37**, 2305–2315 (2007).
25. Van Aken, H. M. *et al.* Observation of a young Agulhas ring, Astrid, during the MARE, the Mixing of Agulhas Rings Experiment, in March 2000. *Deep-Sea Res. II* **50**, 167–195 (2003).
26. Madec, G. *NEMO Ocean Engine* (Note du Pôle de Modélisation, Institut Pierre-Simon Laplace, 2006).
27. Large, W. G. & Yeager, S. G. *Diurnal to Decadal Global Forcing for Ocean and Sea-Ice Models: the Data Sets and Flux Climatologies* (NCAR Technical Note NCAR/TN-460+STR, National Center for Atmospheric Research, 2004).
28. Gent, P. R. & McWilliams, J. C. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **20**, 150–155 (1990).
29. Hsieh, W. W., Davey, M. K. & Wajsowicz, R. C. The free Kelvin wave in finite-difference numerical models. *J. Phys. Oceanogr.* **13**, 1383–1397 (1983).
30. Rio, M. H., Schaeffer, P., Hernandez, F. & Lemoine, J.-M. in *Gocina: Improving Modelling of Ocean Transport and Climate Prediction in the North Atlantic Region using GOCE Gravimetry* (eds Knudsen, P., Johannessen, J., Gruber, T., Stammer, S & van Dam, T.) 6 pp. (Cahiers du Centre Européen de Géodynamique et de Séismologie 25, EGCS, 2005).

LETTERS

Low heat flow inferred from >4 Gyr zircons suggests Hadean plate boundary interactions

Michelle Hopkins[1], T. Mark Harrison[1] & Craig E. Manning[1]

The first ~600 million years of Earth history (the 'Hadean' eon) remain poorly understood, largely because there is no rock record dating from that era. Detrital Hadean igneous zircons from the Jack Hills[1], Western Australia, however, can potentially provide insights into the conditions extant on our planet at that time. Results of geochemical investigations[2–13] using these ancient grains have been interpreted to suggest the presence of a hydrosphere[2–4,7,8] and continental crust[9,10] before 4 Gyr. An underexploited characteristic of the >4 Gyr zircons is their diverse assemblage of mineral inclusions[14–17]. Here we present an examination of over 400 Hadean zircons from Jack Hills, which shows that some inclusion assemblages are conducive to thermobarometry. Our thermobarometric analyses of 4.02–4.19-Gyr-old inclusion-bearing zircons constrain their magmatic formation conditions to about 700 °C and 7 kbar. This result implies a near-surface heat flow of ~75 mW m^{-2}, about three to five times lower than estimates of Hadean global heat flow. As the only site of magmatism on modern Earth that is characterized by heat flow of about one-quarter of the global average is above subduction zones, we suggest that the magmas from which the Jack Hills Hadean zircons crystallized were formed largely in an underthrust environment, perhaps similar to modern convergent margins.

Geochemical and mineralogical investigations of Hadean zircons, including oxygen isotopes[2–4], fission Xe[5,6], zircon thermometry[7,8], Lu–Hf[9,10], Sm–Nd[11,12], trace elements[13] and mineral inclusion studies[14–17], are beginning to reveal insights into the physical and environmental conditions on the early Earth. The ~100 Hadean zircons that have previously been examined for mineral inclusions have been documented to contain quartz, apatite, monazite, K-feldspar, xenotime, rutile, biotite, muscovite, chlorite, FeOOH, Ni-rich pyrite, thorite, amphibole, plagioclase and diamond[14–17]. Two broad inclusion assemblages are consistent with their forming in 'I-type' (for example hornblende, quartz, biotite, plagioclase, apatite, ilmenite) and 'S-type' (quartz, K-spar, muscovite, monazite) granitoids[2]. The

observation of diamond inclusions in Hadean Jack Hills zircons may imply thick, cool lithosphere before 4 Gyr ago[17].

We examined 403 Jack Hills zircons from this period (4.02–4.19 Gyr) for inclusion type and abundance (see Methods). Of these, 85 possessed inclusions that intersected the polished surface. Although only about a quarter of the polished zircon surfaces exposed inclusions, optical examination showed that most grains contain subsurface inclusions. In selected cases we used laser Raman confocal imaging of the zircons embedded in their epoxy mounts but found that epoxy fluorescence could interfere with signals from subsurface phases. Some subsurface minerals, such as titanite, do yield clearly identifiable spectra.

The type and abundance of inclusions identified by energy-dispersive spectroscopy and optical imaging methods are as follows (number, average size, size range): 37% muscovite (31, ~4 μm, 1–10 μm); 36% quartz (30, ~13 μm, 4–25 μm); 12% biotite (9, ~5 μm, 2–10 μm); 7% apatite (5, ~4 μm, 2–10 μm), 2% hornblende (2, 2 and 10 μm), 2% rare-earth element oxide (2, 4 and 25 μm), 1% monazite (1, ~10 μm); 1% albite (1, ~4 μm); 1% ilmenite (1, ~25 μm). Although we have examined over ten times as many Hadean zircons as Menneken et al.[17], we have not identified any diamond inclusions. We obtained chemical analyses on inclusions that were both relevant to barometric methods and large enough to obtain reliable results (that is, >5 μm). Only seven zircons met these criteria: six containing muscovite (+ other phases) and one containing hornblende (+ biotite + titanite).

Six muscovites associated with minerals including quartz, biotite, rutile and feldspar were analysed by electron microprobe analyser (EMPA) for silicon content and fall into two broad groups characterized by Si$_{pfu}$ (per formula unit normalized to 11 oxygens) values of about 3.12 (Group I) and 3.4 (Group II). The inclusions are generally small but yield results consistent with muscovite structural formula (Table 1). Sodium loss seems to be small. The high TiO$_2$ contents

Table 1 | Electron microprobe results for muscovite and hornblende inclusions in Hadean Jack Hills zircons

|	Muscovite						Hornblende
	RSES55_6.15	RSES67_19.5	RSES67_15.16	RSES77_5.7	RSES67_3.2	RSES61_10.8	RSES58_7.9
SiO$_2$	47.5	46.67	45.77	46.46	50.62	51.20	42.16
TiO$_2$	0.4	0.19	0.40	0.40	1.52	0.22	0.46
Al$_2$O$_3$	35.69	34.55	34.14	34.84	25.36	26.30	11.97
MgO	0.92	0.93	0.71	1.11	1.33	1.68	8.08
CaO	0.05	0.05	0.03	0.00	1.29	0.08	11.86
MnO	0.01	0.00	0.00	0.02	0.00	0.05	0.65
∑Fe as FeO	1.73	1.52	1.02	1.30	4.12	6.37	21.53
Na$_2$O	0.45	0.42	0.50	0.36	1.29	0.04	0.71
K$_2$O	9.51	9.49	9.92	9.80	9.69	8.44	0.83
Total	96.26	93.82	92.50	94.28	95.20	94.39	98.20
Si$_{pfu}$	3.11	3.13	3.12	3.11	3.42	3.46	

Si$_{pfu}$ is normalized to 11 oxygens.

[1]Institute of Geophysics and Planetary Physics and Department of Earth and Space Sciences, University of California, Los Angeles, California 90095, USA.

(typically ≥0.4%) are characteristic of derivation from peraluminous granites containing Ti-bearing phases[18]. One hornblende inclusion yields 2.25 Al per 13 cations (Table 1; Fig. 1m,n). Titanium concentrations for the seven zircons, measured by ion microprobe[8] (only analytical errors quoted), range from 3 to 9 p.p.m., yielding apparent crystallization temperatures[7] (T_{zir}) of 665–745 °C (assuming a rutile activity (a_{TiO_2}) of 1). None of the analysed inclusions (Fig. 1) are associated with imperfections in the zircon (cracks, growth zone boundaries). Because of this, and the fact that included white micas were not affected by the widespread alteration of muscovite to fuchsite (a Cr-mica) in the host-rock matrix[1], we conclude that the analysed inclusions remained closed chemical systems since deposition. Details of inclusion mineralogy, zircon age and Ti temperatures are given in Fig. 1.

Experimental results in the system K-feldspar + phlogopite + quartz indicate a broadly linear increase of Si content in white mica (that is, the celadonite component) with pressure[19]. Thermodynamic parameters derived from experimental studies[19] are incorporated into THERMOCALC[20] version 3.26 (2007), permitting construction of pseudo-sections for relevant rock compositions. Using the Bullenbalong cordierite granite as a representative peraluminous magma composition[21], we have calculated Si_{pfu} values as a function of temperature and pressure using a Na-free solution model[22], modified to include ideal mixing of Fe and Mg on octahedral sites. The average values of $T_{zir} = 690 \pm 15$ °C and $Si_{pfu} = 3.12 \pm 0.01$ of the four Group I muscovites indicate formation at $P = 6.9 \pm 0.8$ kbar. Incorporation of Na mixing for K in muscovite[22] yields pressure about 1 kbar higher; TiO_2 and Fe_2O_3 are not accounted for but are negligible (Table 1). Thus, we interpret our result as a minimum estimate of pressure. We found that varying the bulk composition of peraluminous silicic melts had little effect on these calculations provided Na/K was low (that is, muscovite composition was controlled by pressure, temperature, and solid-solution energetics, not bulk composition). This is potentially important as low-degree partial melts from amphibolites yield trondhjemitic[23] magmas with Na/K that is too high to stabilize muscovite on the liquidus, offering further evidence in support of a metasedimentary magma source.

Calculations based on subsolidus experiments suggest that the higher Si_{pfu} values of ~3.4 in Group II muscovites (Table 1) could correspond to pressures of over 20 kbar. However, these calculations are more indicative than conclusive as thermodynamic data for these

Figure 1 | Cathodoluminescence and secondary electron micrographs of inclusion-bearing Jack Hills zircons. Sample identification ($^{207}Pb/^{206}Pb$ age, percentage discordance, Ti-in-zircon crystallization temperature) is as follows. **a, b**, RSES55_6.15 (4,017 ± 19 Myr, −7% discordant, $T_{zir} = 695 \pm 15$ °C), muscovite $Si_{pfu} = 3.11$. Only muscovite inclusions were seen. **c, d**, RSES67_19.5 (4,151 ± 5 Myr, 3% discordant, $T_{zir} = 665 \pm 15$ °C). This zircon contains inclusions of quartz, muscovite and biotite ranging in size from 5 to 8 μm. **e, f**, RSES67_15.16 (4,192 ± 7 Myr, −3% discordant, $T_{zir} = 723 \pm 15$ °C), muscovite $Si_{pfu} = 3.12$. Both muscovite and biotite are present and range in size from 4 to 10 μm. **g, h**, RSES77_5.7 (4,061 ± 14 Myr, 1% discordant, $T_{zir} = 667 \pm 15$ °C), muscovite $Si_{pfu} = 3.11$. This zircon contains a large inclusion with coexisting quartz, rutile (that is, $a_{TiO_2} \approx 1$), and muscovite ranging in size from 10 to 80 μm. The calculated temperature of ~670 °C in a zircon coexisting with rutile is evidence of crystallization from a water-saturated magma[8]. The TiO_2 content (~0.4%) is similar to two of the other three inclusions with muscovite $Si_{pfu} \approx 3.12$ (Table 1) suggesting that they too formed under conditions close to rutile saturation. **i, j**, High-Fe muscovites RSES67_3.2 (coexisting quartz, biotite and muscovite; 4,008 ± 6 Myr, −2% discordant, $T_{zir} = 745 \pm 15$ °C). This analysis seems to contain a contaminating Ca-bearing phase. **k, l**, RSES61_10.8 (4,028 ± 11 Myr, 0% discordant, $T_{zir} = 693 \pm 15$ °C) yields muscovite Si_{pfu} values of ~3.4 (Table 1). **m, n**, RSES58_7.9 (4,025 ± 21 Myr, −9% discordant, $T_{zir} = 700 \pm 15$ °C) contains hornblende with 2.2 Al per 13 cations, indicating an apparent pressure of 7 ± 2 kbar. Although quartz was not directly seen, coexisting biotite and titanite imply quartz saturation in the host melt. musc, muscavite; qtz, quartz; bi, biotite; hbl, hornblende.

494
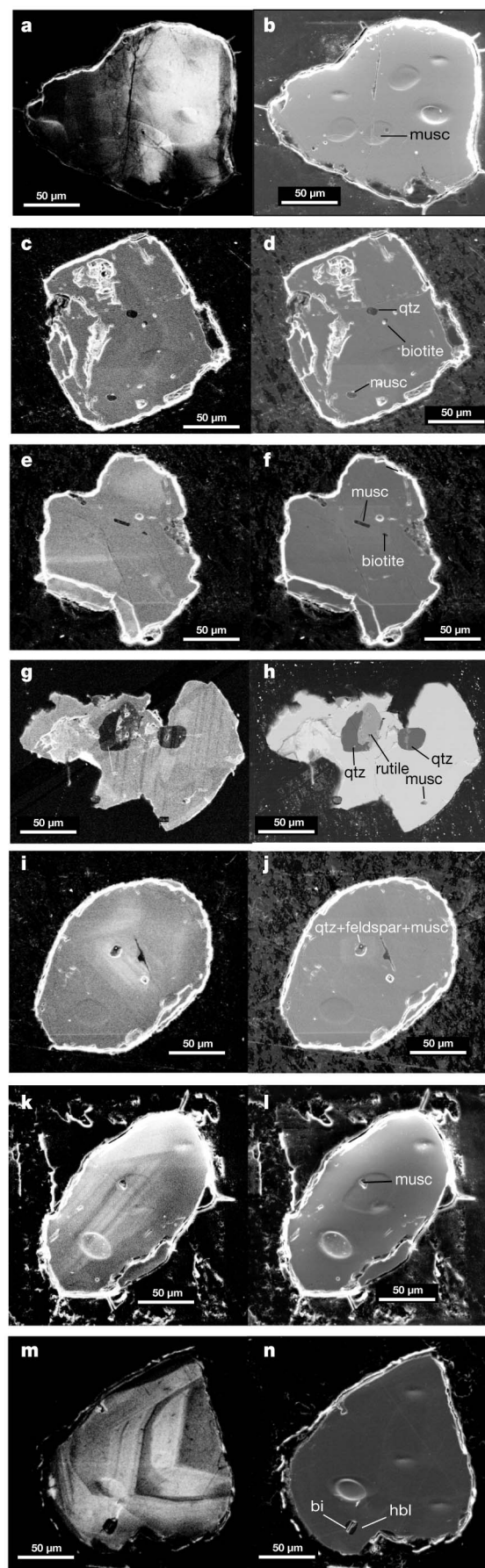
conditions are incomplete (for example, Fe treated as ferrous rather than ferric state).

The empirically calibrated Al-in-hornblende barometer[18] for low-variance granitoids seems to be valid provided that quartz is present and Fe/(Fe+Mg) < 0.65. The total aluminium content of the hornblende inclusion (2.25 Al per 13 cations; Table 1), coupled with the Ti-in-zircon temperature of $700 \pm 15\,°C$, suggests a pressure of 7 ± 2 kbar (ref. 18).

Perhaps the most surprising result of our extensive survey is the observation that coequally abundant muscovite and quartz make up nearly three-quarters of all inclusions in Hadean Jack Hills zircons. The limited stability field of muscovite + quartz in peraluminous granitoids[23] alone restricts temperatures to below $800\,°C$ and pressure to more than 4 kbar (Fig. 2).

Ignoring for the moment the two higher-pressure data, the thermobarometric results indicate a formation environment of ~7 kbar and $\sim690\,°C$. For a surface temperature of $0\,°C$ and rock density of $3\,g\,cm^{-3}$, the average geotherm to the site of zircon crystallization would be $\sim30\,K\,km^{-1}$ (Fig. 3). Because this value is very similar to the slope of the Si_{pfu} isopleths (Fig. 2), underestimation of crystallization temperature due to sub-unity a_{TiO_2} (in those cases where rutile is not present) would be almost exactly compensated by lower calculated pressure; that is, the apparent gradient of about $30\,K\,km^{-1}$ is insensitive to changes in T_{zir}. The two muscovite inclusions suggesting even higher pressures (Table 1) imply an even lower geotherm. Although the report of diamonds[17] in Hadean Jack Hills zircons (Fig. 3) might also be indicative of very low heat flow, we believe that the lack of a calibrated solution model for muscovite under such conditions makes it premature to speculate on their significance.

Figure 3 | Pressure–temperature diagram. The position of the pseudo-section from Fig. 2 is shown in grey relative to approximate modern geotherms, the 'wet' granite solidus and the graphite–diamond stability boundary.

Note that both a Hadean surface temperature of $>0\,°C$ and mechanisms associated with granitoid plutonism (for example crystallization following buoyant magma ascent) would lead us to overestimate the average geotherm to the site of melting. In contrast, we have not identified a plausible mechanism that would have the opposite effect, and thus regard $30\,K\,km^{-1}$ as an upper bound. For a typical rock conductivity[24] of $2.5\,W\,m^{-1}\,K^{-1}$, $30\,K\,km^{-1}$ translates into a near-surface ($\leq40\,km$) heat flow of $\sim75\,mW\,m^{-2}$. This value, slightly lower than the Earth's average heat loss today[25], is substantially less than that inferred for global heat flow during both the Archean[26,27] ($150–200\,mW\,m^{-2}$) and Hadean[28,29] ($160–400\,mW\,m^{-2}$).

As radioactive heat generation was about three times as great at 4.1 Gyr ago as at present[24] and the Earth is generally thought to have cooled by $50–100\,K\,Gyr^{-1}$ (ref. 30), it is difficult to conceive that Hadean global heat flow was less than about three times higher than our upper bound of $\sim75\,mW\,m^{-2}$. Today, the only magmatic environment characterized by heat flow of around 25–30% of the global average is where subducting oceanic lithosphere refrigerates the overlying wedge as it descends into the mantle[25,31] (note that although volcanic arcs have greater near-surface heat flow because of the advection of magmatic heat, the average geotherm to the site of magmagenesis is about $15\,K\,km^{-1}$). Given that the inclusion mineralogy of ancient Jack Hills zircons points towards their origin from hydrous, SiO_2-saturated, meta- and peraluminous melts similar to the two distinctive type of convergent margin magmas observed today (arc-type andesites and Himalayan-type leucogranites), we interpret our results as evidence that many of the Jack Hills Hadean zircons crystallized from magmas generated in an underthrust environment, possibly similar to modern convergent margins.

Figure 2 | Pressure–temperature pseudo-section for model S-type granite. Phase relations in the Bullenbalong cordierite granite[25] in the presence of q, pl and bio (solution models are from Holland and Powell[20] and White *et al.*[32]). Shading becomes lighter with decreasing variance. Phase regions depicted with grey shading denote muscovite unstable with melt, whereas blue colours show regions where muscovite and melt stably coexist. Selected contours of predicted Si content in muscovite (3.07–3.17, in moles per formula unit) are shown with light lines. The highlighted polygon gives the pressure calculated from combination of Si content of muscovite (3.12 ± 0.01) with the temperature from Ti in zircon ($690 \pm 15\,°C$). Phase abbreviations: bi, biotite; mu, muscovite; sil, sillimanite; q, quartz; pl, plagioclase; liq, granitic melt; g, garnet; cd, cordierite; ksp, K feldspar. The error box represents one standard deviation in both temperature and pressure.

METHODS SUMMARY

Sample preparation. We hand-picked zircon grains from heavy mineral concentrates, placed them on double-sided adhesive tape, and mounted them in 1-inch-diameter epoxy disks together with AS-3 standards[2,4]. The disks were then ground with SiC paper and polished only with $1\,\mu m$ diamond paste.

Ion microprobe. The epoxy mounts were gold-coated and surveyed for $^{207}Pb/^{206}Pb$ age using high-resolution ion microprobes at the Australian National University and the University of California, Los Angeles. Precise U-Th-Pb ages were then determined on zircons found to be more than about

4 Gyr old. Titanium and rare-earth element measurements were undertaken using methods described previously[4].

Electron imaging. Mounts with zircons >4 Gyr old were cleaned of their gold coatings and carbon-coated. Samples were imaged in a LEO 1430VP with secondary and backscattered electrons and using cathodoluminescence to identify inclusions and provide information on internal structure of the zircon. We qualitatively chemically characterized inclusions by using energy-dispersive spectroscopy.

Electron microprobe. We obtained quantitative chemical compositions of inclusions *in situ* using a JEOL 8200 electron microprobe equipped with five wavelength-dispersive spectrometers. An electron beam $\sim 1\,\mu m$ in diameter was used with an accelerating voltage of 15 kV. Standard corrections were used.

1. Spaggiari, C. V., Pidgeon, R. T. & Wilde, S. A. The Jack Hills greenstone belt, Western Australia. Part 2: Lithological relationships and implications for the deposition of ≥4.0 Ga detrital zircons. *Precambr. Res.* **155,** 261–286 (2007).
2. Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409,** 178–181 (2001).
3. Cavosie, A. J., Valley, J. W. & Wilde, S. A. Magmatic $\delta^{18}O$ in 4400–3900 Ma detrital zircons: a record of the alteration and recycling of crust in the Early Archean. *Earth Planet. Sci. Lett.* **235,** 663–681 (2005).
4. Trail, D. *et al.* Constraints on Hadean zircon protoliths from oxygen isotopes, REEs and Ti-thermometry. *Geochem. Geophys. Geosyst.* **8,** Q06014 10.1029/2006GC001449 (2007).
5. Turner, G., Harrison, T. M., Holland, G., Mojzsis, S. J. & Gilmour, J. Extinct Pu-244 in ancient zircons. *Science* **306,** 89–91 (2004).
6. Turner, G. *et al.* Pu–Xe, U–Xe, U–Pb chronology and isotope systematics of ancient zircons from Western Australia. *Earth Planet. Sci. Lett.* **261,** 491–499 (2007).
7. Watson, E. B. & Harrison, T. M. Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science* **308,** 841–844 (2005).
8. Harrison, T. M. & Schmitt, A. K. High sensitivity mapping of Ti distributions in Hadean zircons. *Earth Planet. Sci. Lett.* **261,** 9–19 (2007).
9. Harrison, T. M., Schmitt, A. K., McCulloch, M. T. & Lovera, O. M. Early (≥4.5 Ga) formation of terrestrial crust: Lu–Hf, $\delta^{18}O$, and Ti thermometry results for Hadean zircons. *Earth Planet. Sci. Lett.* **268,** 476–486 (2008).
10. Harrison, T. M. *et al.* Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* **310,** 1947–1950 (2005).
11. Amelin, Y. Sm–Nd systematics of zircon. *Chem. Geol.* **211,** 375–387 (2005).
12. Caro, G. *et al.* Precise analysis of $^{142}Nd/^{144}Nd$ in small samples: Application to Hadean zircons from Jack Hills (W. Australia) and diamond inclusions from Finsch (S. Africa). *Chem. Geol.* **247,** 253–265 (2008).
13. Maas, R. & McCulloch, M. T. The provenance of Archean clastic metasediments in the Narryer gneiss complex, Western Australia: Trace element geochemistry, Nd isotopes, and U–Pb ages from detrital zircons. *Geochim. Cosmochim. Acta* **55,** 1915–1932 (1991).
14. Maas, R., Kinny, P. D., Williams, I. S., Froude, D. O. & Compston, W. The Earth's oldest known crust: A geochronological and geochemical study of 3900–4200 Ma old detrital zircons from Mt Narryer and Jack Hills, Western Australia. *Geochim. Cosmochim. Acta* **56,** 1281–1300 (1992).
15. Trail, D., Mojzsis, S. J. & Harrison, T. M. Inclusion mineralogy of pre-4.0 Ga zircons from Jack Hills, Western Australia: A progress report. *Geochim. Cosmochim. Acta* **68,** A743 (2004).
16. Cavosie, A. J., Wilde, S. A., Liu, D., Weiblen, P. & Valley, J. W. Internal zoning and U–Th–Pb chemistry of Jack Hills detrital zircons: a mineral record of early Archean to Mesoproterozoic (4348–1576 Ma) magmatism. *Precambr. Res.* **134,** 251–279 (2004).
17. Menneken, M., Nemchin, A. A., Geisler, T., Pidgeon, R. T. & Wilde, S. A. Hadean diamonds in zircon from Jack Hills, Western Australia. *Nature* **448,** 917–920 (2007).
18. Anderson, J. L. Status of thermobarometry in granitic batholiths. *Trans. R. Soc. Edinb. Earth Sci.* **87,** 125–138 (1996).
19. Massonne, H.-J. & Szpurka, Z. Thermodynamic properties of white micas on the basis of high-pressure experiments in the systems K_2O-MgO-Al_2O_3-SiO_2-H_2O and K_2O-FeO-Al_2O_3-SiO_2-H_2O. *Lithos* **41,** 229–250 (1997).
20. Holland, T. J. B. & Powell, R. An internally consistent thermodynamic data set for phases of petrologic interest. *J. Metamorph. Geol.* **16,** 309–343 (1998).
21. McLaren, S., Sandiford, M., Powell, R., Neumann, N. & Woodhead, J. Palaeozoic intraplate crustal anatexis in the Mount Painter Province, South Australia: Timing, thermal budgets and the role of crustal heat production. *J. Petrol.* **47,** 2281–2302 (2006).
22. Coggon, R. & Holland, T. J. B. Mixing properties of phengitic micas and revised garnet-phengite thermobarometers. *J. Metamorph. Geol.* **20,** 683–696 (2002).
23. Helz, R. T. Phase relations of basalts in their melting ranges at P_{H2O} = 5 kb. Part II. Melt composition. *J. Petrol.* **17,** 139–193 (1976).
24. Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* 2nd edn, 133 (Wiley, 2002).
25. Pollack, H. N., Hurter, S. J. & Johnson, J. R. Heat flow from the Earth's interior: analysis of the global data set. *Rev. Geophys.* **31,** 267–280 (1993).
26. Bickle, M. J. Heat loss from the Earth: Constraints on Archaean tectonics from the relation between geothermal gradients and the rate of plate production. *Earth Planet. Sci. Lett.* **40,** 301–315 (1978).
27. Abbott, D. H. & Hoffman, S. E. Archaean plate tectonics revisited. Part 1. Heat flow, spreading rate, and the age of subducting oceanic lithosphere and their effects on the origin and evolution of continents. *Tectonics* **3,** 429–448 (1984).
28. Smith, J. V. The first 800 million years of Earth's history. *Phil. Trans. R. Soc. Lond. A* **301,** 401–422 (1981).
29. Sleep, N. L. Evolution of the mode of convection within terrestrial planets. *J. Geophys. Res.* **105,** 17563–17578 (2000).
30. Bedini, R. M., Blichert-Toft, J., Boyet, M. & Albarede, F. Isotopic constraints on the cooling of the continental lithosphere. *Earth Planet. Sci. Lett.* **223,** 99–111 (2004).
31. Blackwell, D. D., Steele, J. L. & Carter, L. S. in *Neotectonics of North America* (eds Slemmons, D. B., Engdahl, E. R. and Blackwell, D. D.) 423–437 (DNAG Decade Map 1, Geological Society of America, 1991).
32. White, R. W., Powell, R. W. & Holland, T. J. B. Calculation of partial melting equilibria in the system Na_2O-CaO-K_2O-FeO-MgO-Al_2O_3-SiO_2-H_2O (NCKFMASH). *J. Metamorph. Geol.* **19,** 139–153 (2001).

An ancestral turtle from the Late Triassic of southwestern China

Chun Li[1], Xiao-Chun Wu[2], Olivier Rieppel[3], Li-Ting Wang[4] & Li-Jun Zhao[5]

The origin of the turtle body plan remains one of the great mysteries of reptile evolution. The anatomy of turtles is highly derived, which renders it difficult to establish the relationships of turtles with other groups of reptiles. The oldest known turtle, *Proganochelys* from the Late Triassic period of Germany[1], has a fully formed shell and offers no clue as to its origin. Here we describe a new 220-million-year-old turtle from China, somewhat older than *Proganochelys*, that documents an intermediate step in the evolution of the shell and associated structures. A ventral plastron is fully developed, but the dorsal carapace consists of neural plates only. The dorsal ribs are expanded, and osteoderms are absent. The new species shows that the plastron evolved before the carapace and that the first step of carapace formation is the ossification of the neural plates coupled with a broadening of the ribs. This corresponds to early embryonic stages of carapace formation in extant turtles, and shows that the turtle shell is not derived from a fusion of osteoderms. Phylogenetic analysis places the new species basal to all known turtles, fossil and extant. The marine deposits that yielded the fossils indicate that this primitive turtle inhabited marginal areas of the sea or river deltas.

New fossil turtles have been collected from the lower Carnian[2,3] near Guanling in Guizhou Province, China. The sediments were deposited in the Nanpanjiang Trough Basin, which during the early to middle Carnian marine transgression remained surrounded by the Sichuan–Yunnan–Guizhou Old Land on three sides, opening into the Palaeotethys towards the southwest[3,4]. Sedimentological clues indicate anoxic conditions for the bottom water, and driftwood as well as other plant remains indicate the relative proximity of coastal waters[3]. Vast outcrops yielded a rich fauna of invertebrates and marine reptiles[3,5], but turtles from the Guanling Biota are reported here for the first time.

Order Testudinata
New Family Odontochelyidae
Odontochelys gen. nov.

Etymology. '*odonto-*' (comb. form of '*odoús*', Greek): tooth; *chelys* (Greek): lyre of ancient Greece, or tortoise.
Type species. *Odontochelys semitestacea* sp. nov.
Distribution. Late Triassic, Guizhou Province, China.
Diagnosis. As for species.

Odontochelys semitestacea sp. nov.

Etymology. *testaceus* (-*um*) (Latin): of animals, covered with shell; semi- (Latin, as prefix): half.
Holotype. IVPP (Institute of Vertebrate Paleontology and Paleoanthropology) V 15639.
Paratype. IVPP V 13240.
Referred specimen. IVPP V 15653.
Location. Guanling, Guizhou Province, southwestern China.
Horizon. Wayao Member of the Falang Formation.

Diagnosis. A turtle of relatively small size; teeth present on upper and lower jaws, and on kinetic dermal palate; snout elongate and pointed; plastron elongated oval, with four frilled lateral spines; epiplastral dorsal process present; two pairs of mesoplastra; hypoischium butterfly-shaped, closing anal opening behind ischium; only neural plates ossified in carapace; dorsal ribs broadened.

The skull of *Odontochelys* shows an elongate preorbital region forming a pointed snout (Fig. 1a, b). Small and peg-like, pointed teeth are present on premaxilla, maxilla, dentary, pterygoid and vomer (Figs 1b, c and 2c, d). The jugal shows a rounded posteroventral margin, indicating a weak ventral excavation of the cheek. The skull table is not excavated posteriorly. The temporal region is not fenestrated. Distinct basipterygoid processes form an open basicranial articulation with the pterygoids. The pterygoid shows a distinct transverse process that may have separated a subtemporal fenestra from a suborbital fossa (Fig. 2c, d). The quadrate is weakly concave posteriorly; a cavum tympani is absent.

All vertebrae are weakly amphicoelous. The vertebral count is eight cervicals, nine dorsals, two sacrals, and minimally 20 caudals. The cervical centra are distinctly keeled ventrally. Cervical ribs are small, knob-like structures. Eight dorsal ribs are distinctly broadened, single-headed, and articulate in facets located at the middle of the centrum. The last dorsal rib is slender. The distally expanded sacral ribs are not fused with the sacral vertebrae (Fig. 1d). The caudal transverse processes likewise are free. Chevrons articulate on the posteroventral aspect of the preceding centrum (Fig. 2a). A tail-club is absent.

The epiplastral dorsal process shows no trace of a suture at its base. The scapular dorsal blade is rod-like, without an acromial process (Figs 1d and 3c). The broad, curved and plate-like coracoid shows a foramen with its anterior margin open (Fig. 3c). The humerus is more massively built than the femur, but of equal length. The proximal end of the humerus shows well-developed medial and lateral processes, separated by the intertubercular fossa (Fig. 2a, b). The distal expansion of the humerus carries double articulations for the radius and ulna. The radius is slightly shorter and more lightly built than the ulna; the ulna lacks an ossified olecranon. Eight to nine ossifications are present in the carpus: a radiale seems to be fused with the intermedium (in the paratype, absent in the holotype), ulnare, pisiforme, lateral and medial centralia, and five distal carpals (Fig. 2a, b). The metacarpals are proximally overlapping, the third and the fourth being the longest in the series. The phalangeal formula in the manus is 2-3-4-4-3.

The ilium is stout, with a short dorsal shaft (Fig. 1d). Pubes and ischia form a pubo-ischial plate with a medioventral ridge terminating in a posterior ischial tubercle (Fig. 2a, b); the thyroid foramen is present in the referred specimen (Fig. 3d). The pubis carries a ventrally projecting lateral process that articulates with the plastron (Fig. 3d). The ischium is located behind the plastron, and carries a

[1]Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. [2]Canadian Museum of Nature, PO Box 3443, Station D, Ottawa, Ontario K1P 6P4, Canada. [3]Department of Geology, The Field Museum, 1400 South Lake Shore Drive, Chicago, Illinois 60605-2496, USA. [4]Geological Survey of Guizhou Province, Guiyang 550005, China. [5]Zhejiang Museum of Nature History, Hangzhou 310012, China.
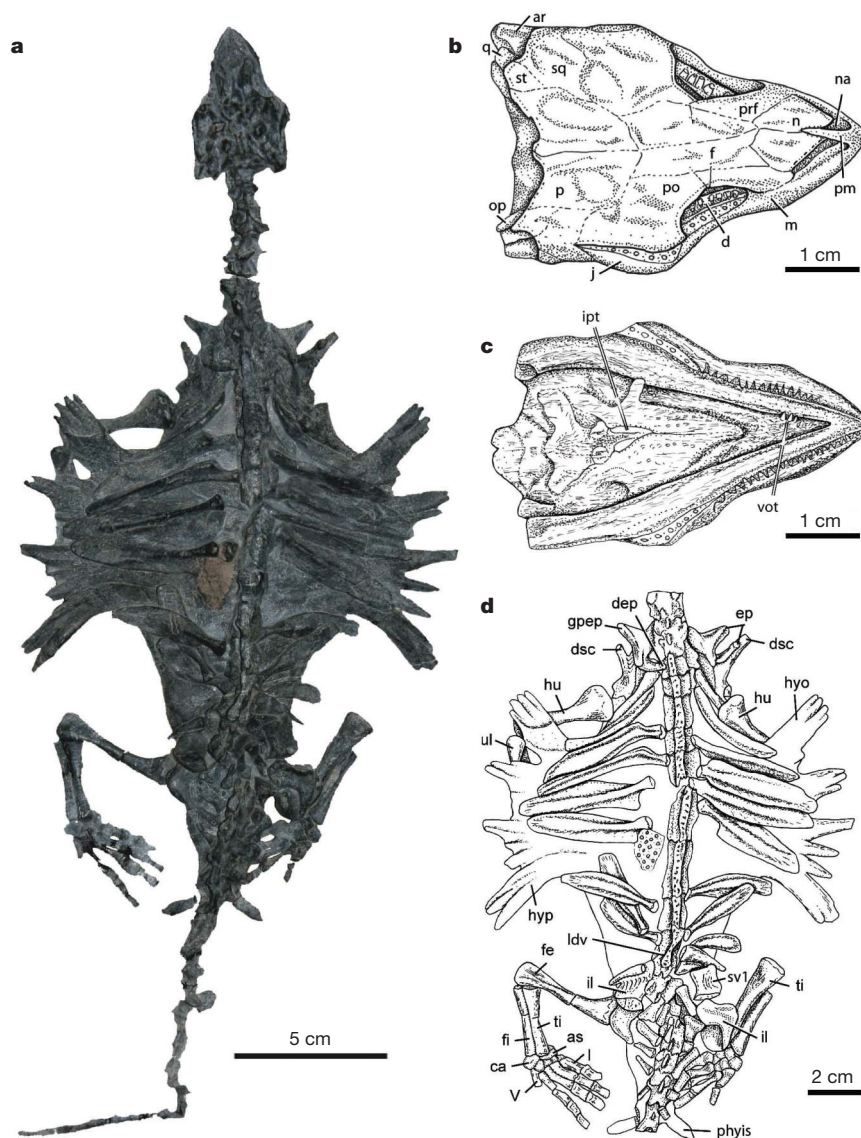
Figure 1 | Holotype (IVPP V 15639) of *Odontochelys semitestacea* gen. et sp. nov. a, Skeleton in dorsal view. **b**, Skull in dorsal view. **c**, Skull in ventral view. **d**, Body in dorsal view. Teeth on the upper jaw and palatal elements were scratched out during excavation. Abbreviations: ar, articular; as, astragalus; ca, calcaneum; d, dentary; dep, dorsal process of epiplastron; dsc, dorsal process of scapula; ep, epiplastron; fe, femur; fi, fibula; gpep, gular projection of epiplastron; hu, humerus; hyo, hyoplastron; hyp, hypoplastron; il, ilium; ipt, interpterygoid vacuity; j, jugal; ldv, last dorsal vertebra; m, maxilla; n, nasal; na, naris; op, opisthotic; p, parietal; phyis, posterolateral process of hypoischium; pm, premaxilla; po, postorbital; prf, prefrontal; q, quadrate; sq, squamosal; st, supratemporal; sv1, 1st sacral vertebra; ti, tibia; ul, ulna; vot, vomerine teeth; I, V, 1st and 5th metatarsals.

posterolaterally projecting ischial process. These processes, together with hypoischium, enclose a rounded anal opening (Fig. 2b). The hypoischium extends into posterolaterally projecting processes. The femur is slender and sigmoidally curved; its proximal articular head is angled away from its long axis. The internal trochanter is distinct. The distal articulations on the femur are confluent. The fibula is more lightly built than the tibia. Astragalus and calcaneum are sutured in the holotype but fused in the paratype; no perforating foramen is located between them (Fig. 2b). No centralia are ossified in the tarsus; four distal tarsals are preserved. The metatarsals are proximally overlapping; the fifth metatarsal is angulated (hooked). The pedal phalangeal formula is 2-3-4-4-?.

Costal and marginal carapacial plates are absent, but neural plates are ossified. The neural plates are not fused with broadened neural spines of dorsal vertebrae but were displaced laterally during fossilization (Fig. 3a, b). The plastron is composed of epiplastra and entoplastron, one pair of hyoplastra, two pairs of mesoplastra, one pair of hypoplastra and one pair of xiphiplastra. The entoplastron is heart-shaped, with well-developed anterolateral facets receiving the epiplastra (Fig. 2b). Hyoplastron, mesoplastron and hypoplastron each bear laterally projecting frilled spines. No osteoderms are present on the neck, trunk, tail and limbs.

Odontochelys is more primitive than *Proganochelys*[1] in presence of teeth on premaxilla, maxilla and dentary; relatively long preorbital skull; distinct transverse process on pterygoid; absence of fully formed carapace; no acromial process on scapula; dorsal ribs articulating at midline of centrum; free sacral ribs; free caudal transverse processes; presence of long tail; four (rather than three) phalanges in digits III and IV of manus and pes; absence of osteoderms and tail-club. *Odontochelys* shares with *Proganochelys* primitive features that are absent in Casichelydia: teeth on vomer and pterygoid; open basicranial articulation; dorsal epiplastral process (also present in
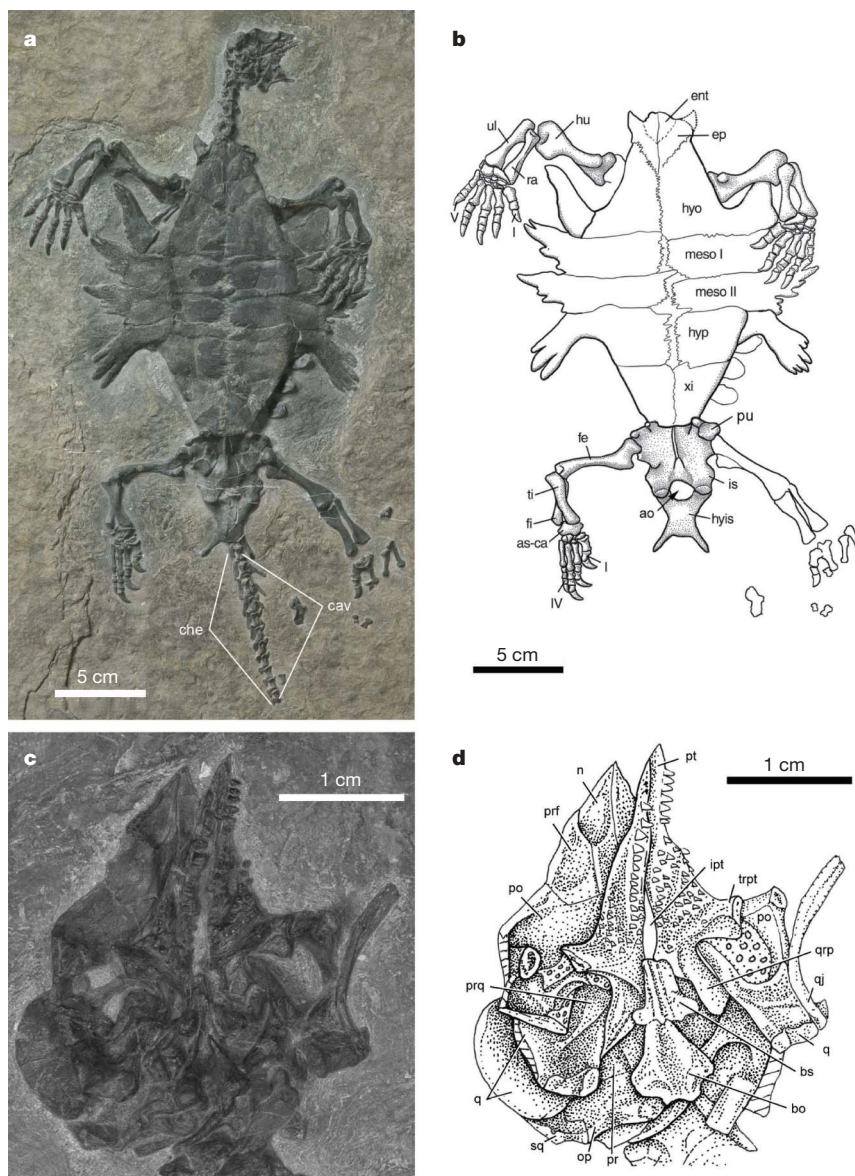
Figure 2 | Paratype (IVPP V 13240) of *Odontochelys semitestacea* gen. et sp. nov. a, Skeleton in ventral view. **b**, Body in ventral view. **c, d**, Skull in ventral and slightly lateral views. Abbreviations as in Fig. 1, plus: ao, anal opening; bo, basioccipital; bs, basisphenoid; cav, caudal vertebrae; che, chevron; ent, entoplastron; hyis, hypoischium; is, ischium; meso I and II, mesoplastra 1 and 2; pr, prootic; prq, pterygoid ramus of quadrate; pt, pterygoid; pu, pubis; qj, quadratojugal; qrp, quadrate ramus of pterygoid; ra, radius; trpt, transverse process of pterygoid; xi, xiphiplastron; I, IV and V, 1st, 4th and 5th digits.

Kayentachelys[6]); broad and plate-like coracoid; ilium with short dorsal shaft; hypoischium present; distinct gular projections on epiplastra. Testing the phylogenetic relationships (see Supplementary Information) of *Odontochelys* confirms its position as basal to all other known turtles, fossil or extant (Fig. 3e). The relationships of turtles with other amniotes have been controversial[1,7–11]. The inclusion of *Odontochelys* in the analysis of turtle relationships within amniotes (see Supplementary Information) supports the position of turtles within diapsid reptiles.

Odontochelys provides documentation that in turtles, the plastron evolved before the carapace. This corresponds to the ossification of plastral before carapacial elements in embryonic turtles[12–14]. The new taxon also shows that the neural plates were the first to evolve among the carapacial elements. The neural plates remained separate from the underlying dorsal neural spines in *Odontochelys*, as was also reported for a juvenile *Testudo loveridgii*[15]. This differs from other turtles[12,16,17], in which neural plate formation expands from the perichondral

ossification of the vertebral neural arch. The dorsal ribs of *Odontochelys* are characteristically broadened, resembling the ribs of *Chelydra serpentina* at the Yntema[18] embryonic stage 23; that is, at a stage before marginal, nuchal and pygal elements start ossification[12,14]. In extant turtles[12–14,16,17,19,20], ossification of costal plates spreads from the perichondral ossification of the dorsal ribs through the adjacent dermis of the carapacial disk. *Odontochelys* resembles embryonic stages of extant turtles in that there is only some broadening and consequent flattening of the dorsal ribs, which do not expand into costal plates. That *Odontochelys* should be represented by juvenile specimens can be rejected, given the fusion of the astragalus and calcaneum. Carapace reduction in fossil and extant turtles[21] never restores the embryonic morphology of the ribs, and is generally correlated with the formation of fontanelles in the plastron that are larger than the narrow midline fontanelle seen in the plastron of *Odontochelys*.

Osteoderms of reptiles form through metaplastic ossification[22–27]. Both dermal sclerification[22] and osteoblast activity contribute to
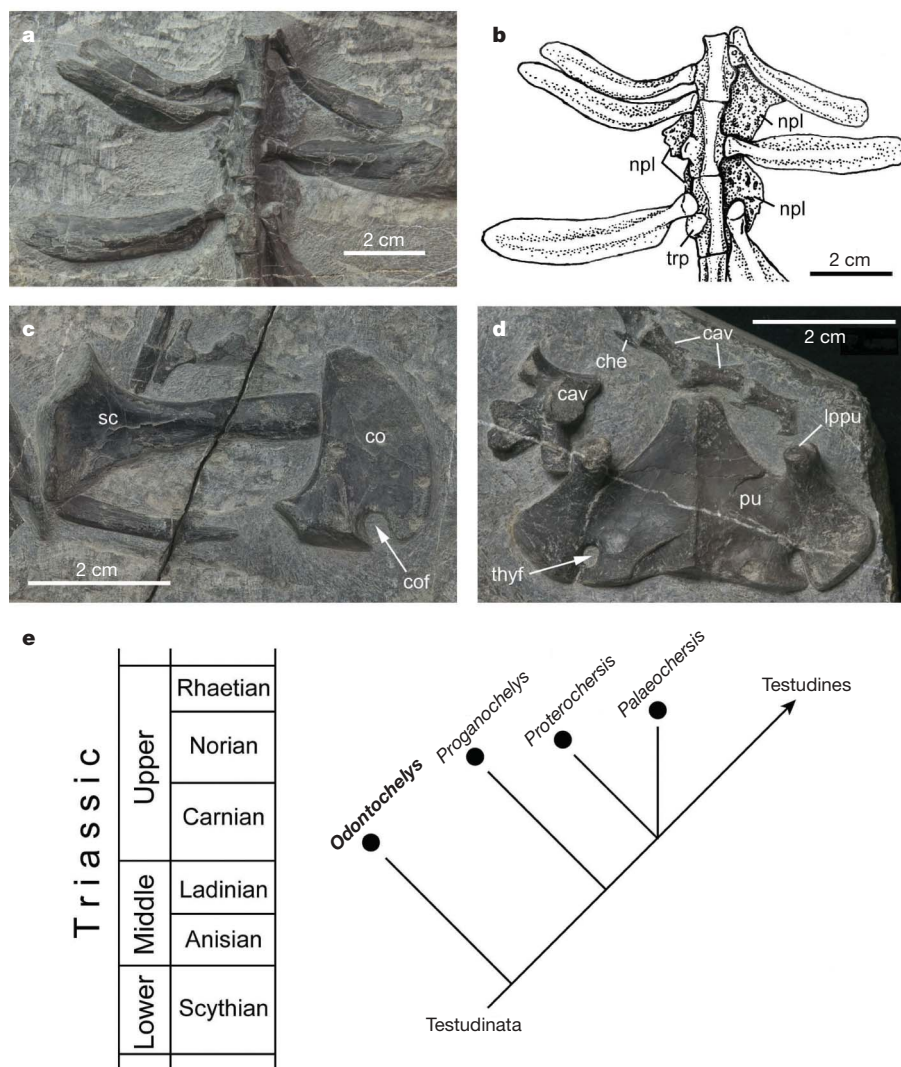
Figure 3 | Referred specimen (IVPP V 15653) of *Odontochelys semitestacea* gen. et sp. nov. a, b, Three dorsal vertebrae, ribs and neural plates (shifted onto one side; others are not exposed) in ventral views. **c,** Scapula and coracoid in medial view. **d,** Two pubes in inner and posterior caudal vertebrae in lateral views. **e,** Geological time table of the Triassic period, showing the stratigraphic distribution of basal members of Testudinata, and a cladogram depicting the phylogenetic relationships of *Odontochelys* with other Triassic turtles. Abbreviations as in Figs 1 and 2, plus: co, coracoid; cof, coracoid foramen; lppu, lateral process of pubis; npl, neural plate; sc, scapula; thyf, thyroid fenestra; trp, transverse process of dorsal vertebrae.

carapace formation in turtles[17,28]. This has led to the hypothesis that turtles evolved from an osteoderm-bearing ancestor[8,17,29]. The evolution of the carapace would thus have involved the shift of endoskeletal neural arches and ribs into the deep dermis of the carapacial disk, followed by fusion of endoskeletal and dermal elements, or a replacement of the former by the latter[17]. The absence of osteoderms in *Odontochelys* indicates that the carapace did not evolve through the fusion of ancestrally present osteoderms[8], but instead formed by intramembranous ossification within the carapacial disk[11,16,19,20].

The evolution of a ventral dermal armor (plastron) has previously been taken to indicate an aquatic origin of turtles[11]. *Odontochelys* may have inhabited marginal areas of the Nanpanjiang Trough Basin, or river delta habitats. Forelimb proportions have been shown to correlate with habitat preferences in living turtles[30]. Studies of primitive fossil turtles such as *Proganochelys* and *Palaeochersis* showed that they were predominantly terrestrial. In addition, shell bone histology[28] was used to support the hypothesis of a terrestrial origin of turtles[30]. However, on the basis of its forelimb proportions, *Odontochelys* compares to living turtles that inhabit 'stagnant or small bodies of water'[30]. This is indicative of primarily aquatic habits and of a possible aquatic origin of turtles.

1. Gaffney, E. S. The comparative osteology of the Triassic turtle *Proganochelys. Bull. Am. Mus. Nat. Hist.* **194,** 1–263 (1990).
2. Wang, L. *et al.* Biostratigraphy of Triassic marine reptiles in southwest Guizhou and its adjacent area. *Acta Geol. Sin.* **75,** 349–353 (2001).
3. Wang, X. *et al.* The Late Triassic black shales from the Guanling area. Guizhou Province, south-west China: a unique marine reptile and pelagic crinoid fossil Lagerstätte. *Palaeontology* **51,** 27–61 (2008).
4. Wang, X., Hagdorn, H. & Wang, C. Pseudoplanktonic lifestyle of the Triassic crinoid *Traumatocrinus* from southwest China. *Lethaia* **39,** 187–193 (2006).
5. Jiang, D. *et al.* Guanling biota: a marker of Triassic biotic recovery from end-Permian extinction in the ancient Guizhou sea. *Acta Geol. Sin.* **79,** 729–738 (2005).
6. Joyce, W. J. Phylogenetic relationships of Mesozoic turtles. *Bull. Peabody Mus. Nat. Hist.* **48,** 3–102 (2007).
7. Reisz, R. R. & Laurin, M. *Owenetta* and the origin of turtles. *Nature* **349,** 324–326 (1991).
8. Lee, M. S. Y. The origin of the turtle body plan: bridging a famous morphological gap. *Science* **261,** 1716–1720 (1993).
9. Lee, M. S. Y. Correlated progression and the origin of turtles. *Nature* **379,** 812–815 (1996).
10. Rieppel, O. & deBraga, M. Turtles as diapsid reptiles. *Nature* **384,** 453–455 (1996).
11. Rieppel, O. & Reisz, R. R. The origin and early evolution of turtles. *Annu. Rev. Ecol. Syst.* **30,** 1–22 (1999).

12. Rieppel, O. Studies on skeleton formation in reptiles. Patterns of ossification in the skeleton of *Chelydra serpentina* Linnaeus (Reptilia, Testudines). *J. Zool.* **231**, 487–509 (1993).

13. Scheil, C. A. Osteology and skeletal development of *Apalone spinifera* (Reptilia: Testudines: Trionychidae). *J. Morphol.* **256**, 42–78 (2003).

14. Sheil, C. A. & Greenbaum, E. Reconsideration of skeletal development of *Chelydra serpentina* (Reptilia: Testudinata: Chelydridae): evidence for intraspecific variation. *J. Zool.* **265**, 235–267 (2005).

15. Procter, J. B. A study of the remarkable tortoise, *Testudo loveridgii* Blgr, and the morphogeny of the chelonian carapace. *Proc. Zool. Soc. Lond.* **1922**, 483–526 (1922).

16. Gilbert, S. F., Loredo, G. A., Brukman, A. & Burke, A. C. Morphogenesis of the turtle shell: the development of a novel structure in tetrapod evolution. *Evol. Dev.* **3**, 47–58 (2001).

17. Scheyer, T. M., Brüllmann, B. & Sánchez-Villagra, M. R. The ontogeny of the shell in side-necked turtles, with emphasis on the homologies of costal and neural bones. *J. Morphol.* **269**, 1008–1021 (2008).

18. Yntema, C. L. A series of stages in the embryonic development of *Chelydra serpentina*. *J. Morphol.* **125**, 219–252 (1968).

19. Kälin, J. Zur Morphogenese des Panzers bei den Schildkröten. *Acta Anat.* **1**, 144–176 (1945).

20. Gilbert, S. F., Cebra-Thomas, J. A. & Burke, A. C. in *Biology of Turtles* (eds Wyneken, J., Godfrey, M. H. & Bels, V.) 1–16 (CRC Press, 2008).

21. Zangerl, R. The homology of the shell elements in turtles. *J. Morphol.* **65**, 383–406 (1939).

22. Moss, M. L. Comparative histology of dermal sclerifications in reptiles. *Acta Anat.* **73**, 510–533 (1969).

23. Zylbernerg, L. & Castanet, J. New data on the structure and growth of the osteoderms in the reptile *Anguis fragilis* L. (Anguidae, Squamata). *J. Morphol.* **186**, 327–342 (1985).

24. Lervrat-Calviac, V. & Zylberg, L. The structure of the osteoderms in the gecko: *Tarentola mauritanica*. *Am. J. Anat.* **176**, 437–446 (1986).

25. Scheyer, T. M. & Sander, P. M. Histology of ankylosaur osteoderms: implications for systematics and function. *J. Vertebr. Paleontol.* **24**, 874–893 (2004).

26. Maine, R. P., de Ricqlès, A., Horner, J. R. & Padian, K. The evolution and function of thyreophoran dinosaur scutes: implications for plate function in stegosaurs. *Paleobiology* **31**, 291–314 (2005).

27. Vickaryous, M. K. & Hall, B. K. Development of the dermal skeleton in *Alligator mississippiensis* (Archosauria, Crocodylia) with comments on the homology of osteoderms. *J. Morphol.* **269**, 398–422 (2008).

28. Scheyer, T. M. & Sander, P. M. Shell bone histology indicates terrestrial paleoecology of basal turtles. *Proc. R. Soc. B* **274**, 1885–1893 (2007).

29. Versluys, J. Über die Phylogenie des Panzers der Schildkröten und über die Verwandtschaft der Lederschildkröte (*Dermochelys coriacea*). *Paläontol. Z.* **1**, 321–347 (1914).

30. Joyce, W. G. & Gauthier, J. A. Paleoecology of Triassic stem turtles sheds new light on turtle origins. *Proc. R. Soc. B* **271**, 1–5 (2003).

LETTERS

Self-renewal and expansion of single transplanted muscle stem cells

Alessandra Sacco[1]*, Regis Doyonnas[1]*†, Peggy Kraft[1], Stefan Vitorovic[1] & Helen M. Blau[1]

Adult muscle satellite cells have a principal role in postnatal skeletal muscle growth and regeneration[1]. Satellite cells reside as quiescent cells underneath the basal lamina that surrounds muscle fibres[2] and respond to damage by giving rise to transient amplifying cells (progenitors) and myoblasts that fuse with myofibres. Recent experiments showed that, in contrast to cultured myoblasts, satellite cells freshly isolated[3–5] or satellite cells derived from the transplantation of one intact myofibre[6] contribute robustly to muscle repair. However, because satellite cells are known to be heterogeneous[4,6,7], clonal analysis is required to demonstrate stem cell function. Here we show that when a single luciferase-expressing muscle stem cell is transplanted into the muscle of mice it is capable of extensive proliferation, contributes to muscle fibres, and Pax7[+]luciferase[+] mononucleated cells can be readily re-isolated, providing evidence of muscle stem cell self-renewal. In addition, we show using in vivo bioluminescence imaging that the dynamics of muscle stem cell behaviour during muscle repair can be followed in a manner not possible using traditional retrospective histological analyses. By imaging luciferase activity, real-time quantitative and kinetic analyses show that donor-derived muscle stem cells proliferate and engraft rapidly after injection until homeostasis is reached. On injury, donor-derived mononucleated cells generate massive waves of cell proliferation. Together, these results show that the progeny of a single luciferase-expressing muscle stem cell can both self-renew and differentiate after transplantation in mice, providing new evidence at the clonal level that self-renewal is an autonomous property of a single adult muscle stem cell.

We reasoned that prospective isolation of muscle stem cells (MuSCs) in conjunction with a dynamic analysis of their fate in vivo would greatly enhance our understanding of their potential to regenerate damaged muscle. Accordingly, we tested various fluorescence-activated cell sorting (FACS) fractionation procedures[3–5,7] and determined that, after depletion of CD45 (also known as Ptprc), CD11b (Itgam), Sca1 (Ly6a) and CD31 (Pecam1), a combination of endogenous markers—CD34 and integrin-α_7—enriched for a muscle cell population of morphologically round cells that uniformly expressed the satellite-cell-specific transcription factor Pax7 (Fig. 1a–c). When isolated from Myf5-nLacZ[Het] mice (in which the nLacZ reporter gene has been introduced into the locus of the myogenic transcription factor Myf5 gene, expression of which is characteristic of activated satellite cells) and plated in vitro, these cells were activated (β-galactosidase[+] (β-gal[+]); ref. 8 and Fig. 1d) and differentiated to form multinucleated myotubes (Fig. 1e). To determine the heterogeneity of this cell population, we performed single-cell reverse transcription followed by polymerase chain reaction for four transcripts for each of 40 FACS-isolated MuSCs (Fig. 1f and Supplementary Methods). Remarkably, all cells (100%) within this population expressed Pax7 and Myf5, characteristic of satellite cells[9,10]. Of these, 25% also expressed MyoD (also known as

Myod1), a marker of committed progenitors[11], and 12% expressed Pax3, a marker of progenitors intermediate between satellite cells and myoblasts in hindlimb muscles[12,13]. These data show that the CD34[+]integrin-α_7[+] population is not fully homogeneous, because a subset of cells expressed commitment markers. To test the function of this muscle cell population in vivo, cells were transplanted into tibialis anterior muscles of non-obese diabetic/severely combined immuno-deficient (NOD/SCID) mice depleted of endogenous satellite cells by 18 Gy irradiation[14,15]. Four weeks after transplantation, muscles were damaged with notexin (NTX)[16,17], and Myf5–β-gal[+] donor-derived cells were subsequently detected in the satellite cell position underneath the basal lamina of myofibres (Fig. 1g, h). This classical histological analysis demonstrated that this population of MuSCs homed to the satellite cell niche and responded appropriately to muscle damage by upregulating expression of the Myf5 transcription factor.

We reasoned that a dynamic assay could complement histological analyses by providing insights into the kinetics and extent of proliferation of transplanted MuSCs. We therefore developed a non-invasive bioluminescence imaging assay to monitor MuSCs by mating Myf5-nLacZ[Het] mice with firefly luciferase (FLuc) transgenic mice[18]. In these studies, cell number was assessed as the bioluminescence signal derived from constitutive luciferase activity, and the activity of the Myf5 promoter was assayed histologically as β-gal activity. The linearity, sensitivity and reproducibility of the bioluminescence assay for quantifying cell numbers was validated in vitro (Supplementary Fig. 1) and in vivo (Fig. 2a). The minimum number of cells detectable above control uninjected legs was 10,000 (Fig. 2a).

To validate bioluminescence imaging as an assay for MuSC function in vivo, we compared freshly isolated integrin-α_7[+]CD34[+] MuSCs with cultured primary myoblasts, because previous studies by others indicated that these two cell types differ markedly[3]. We injected either 5,000 freshly isolated MuSCs or a fourfold excess of 20,000 cultured myoblasts[19–22], both isolated from the Myf5-nLacZ/FLuc transgenic mice, into irradiated legs of NOD/SCID recipients. Four weeks after transplantation, myoblasts were barely detectable (0.2 ± 0.01 × 10^5 photons cm^{-2} s^{-1}; Fig. 2b, top panels), indicating that their numbers had declined, whereas freshly isolated MuSCs yielded robust luciferase activity (29.0 ± 7.0 × 10^5 photons cm^{-2} s^{-1}), a signal corresponding to ~3 × 10^5 cells (Fig. 2b, top panels), which is approximately a 60-fold expansion (~6 doublings). Histological analysis revealed luciferase[+] myofibres in muscles of mice injected with freshly isolated MuSCs, but not myoblasts (Fig. 2b, middle panels). Histochemistry of NTX-damaged muscles revealed the presence of Myf5–β-gal[+] cells, indicative of activated satellite cells, after injection of uncultured MuSCs, but not myoblasts (Fig. 2b, bottom panels). Together, these results confirm that freshly isolated MuSCs, but not myoblasts, successfully engraft, proliferate and give rise to committed progenitors that contribute to muscle fibres.

---

[1]Baxter Laboratory in Genetic Pharmacology, Department of Microbiology and Immunology, Stem Cell Institute, Stanford University School of Medicine, Stanford, California 94305-5175, USA. †Present address: Pfizer Global Research & Development, Genetically Modified Models Center of Emphasis, Groton, Connecticut 06340, USA.
*These authors contributed equally to this work.

To determine the proportion of cells with engraftment potential in this muscle cell population, we transplanted different numbers of freshly isolated MuSCs into irradiated tibialis anterior muscles. Bioluminescence was assayed four weeks after transplantation, and successful engraftment was defined as persistence of a signal >20,000 photons $cm^{-2} s^{-1}$, significantly above the background signal detected in control uninjected legs (Fig. 2c). More than 80% of mice showed engraftment when high numbers of MuSCs (500–5,000) were transplanted; however, even when as few as 10 cells were transplanted, 16% (2 out of 12 mice) showed engraftment (Fig. 2c). This percentage is probably the result of several hurdles such as the heterogeneity of the cell population (Fig. 1f), the survival rate of the cells after the isolation and injection procedures, and the threshold of detection by bioluminescence imaging. Notably, the signal plateaued in all cases (Fig. 2d), as reported for haematopoiesis[23], presumably

reflecting a proliferation of cells until the need is met, after which a combination of cell death and quiescence leads to tissue homeostasis. The plateau occurred later and at a lower level when fewer cells were injected, suggesting that over time endogenous radiation-resistant satellite cells[15] increasingly compete with transplanted MuSCs for engraftment. In contrast with the marked increase in cell numbers that accompanies engraftment of freshly isolated MuSCs, cultured myoblasts decline in numbers within the first few days and then persist without detectable proliferation (Fig. 2e), consistent with previously published reports[24]. Furthermore, when MuSC-transplanted muscles with different engraftment levels were analysed, a linear correlation between the number of donor-derived myofibres and the bioluminescence signal was observed (Fig. 2f), further validating this dynamic assay as an ideal tool to quantify donor-derived cells *in vivo*.
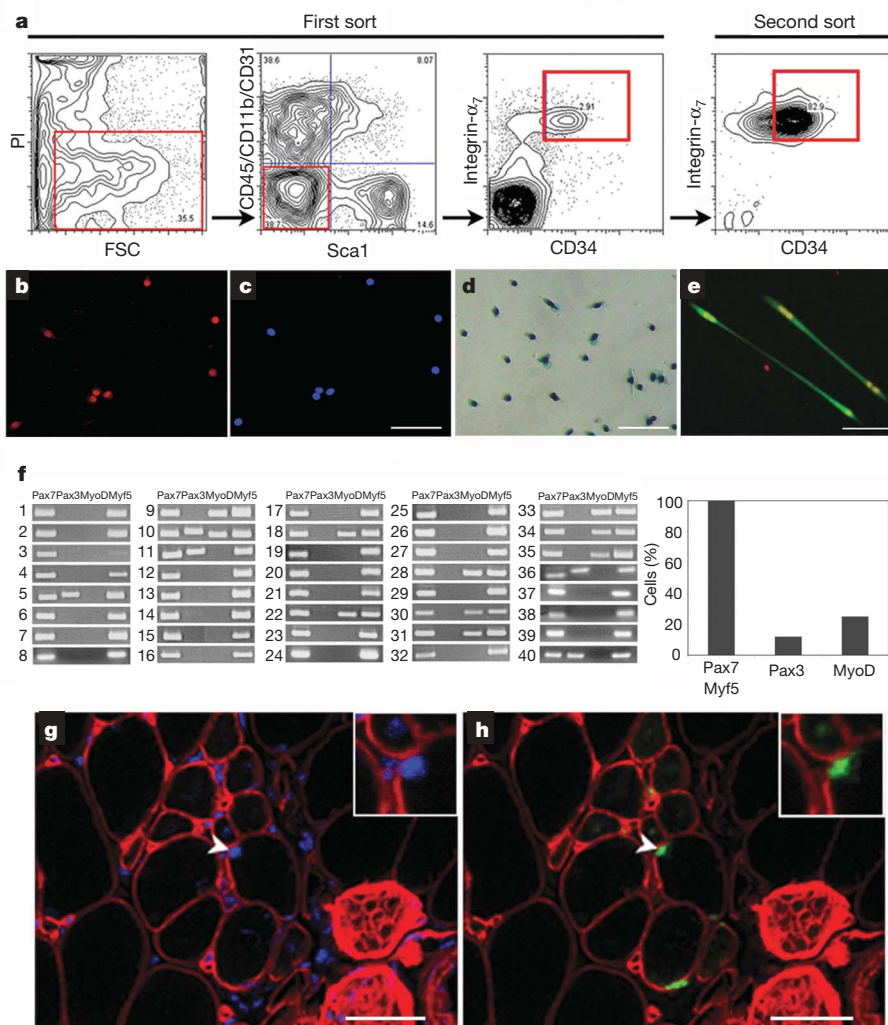
Figure 1 | Characterization of the integrin-α_7^+CD34$^+$ fraction as muscle stem cells. a, Flow cytometry analysis of freshly isolated muscle cells. Living cells were gated for forward scatter (FSC) and propidium iodide (PI) negativity (left panel). Cells negative for blood markers CD45 and CD11b, for endothelial marker CD31 and for mesenchymal marker Sca1 were gated (middle-left panel), and within this population the integrin-α_7^+CD34$^+$ fraction was sorted (right panels). **b, c,** Forty-eight hours after isolation, muscle integrin-α_7^+CD34$^+$ cells were stained for Pax7 (left: Pax7, red; right: nuclei, blue). Scale bar, 80 μm. **d,** After 5 days of culture in growth medium, integrin-α_7^+CD34$^+$ cells isolated from *Myf5-nLacZ^{Het}* mice showed β-gal activity. Scale bar, 80 μm. **e,** After 3 days of culture in differentiation medium, integrin-α_7^+CD34$^+$ cells (from *GFP*-transgenic mice) differentiated to form myotubes (green) that expressed myogenin (red).

Scale bar, 80 μm. **f,** Single MuSCs (1–40) were individually sorted and reverse transcription followed by polymerase chain reaction performed as described in Supplementary Methods. The results show that this population consistently and homogeneously expresses Pax7 and Myf5, the expected transcriptional profile for satellite cells[8–10]. In contrast, both MyoD and Pax3 expression are heterogeneous, indicating the presence of committed progenitors, in accordance with previously published reports[11–13]. **g, h,** Freshly isolated integrin-α_7^+CD34$^+$ cells from *Myf5-nLacZ^{Het}* mice were transplanted into recipient mice. One month after transplant, recipient muscles were damaged by NTX injection and 5 days later immunofluorescence of transverse tissue sections revealed Myf5–β-gal$^+$ cells engrafted in the satellite cell position (arrowhead and inset). β-gal, green; laminin, red; nuclei, blue. Scale bars, 20 μm.

A functional property of adult stem cells is the ability to repeatedly respond to tissue injury by giving rise to substantial numbers of proliferative progenitors. Accordingly, irradiated NOD/SCID mice were injected with 10 or 500 MuSCs (Fig. 3 and Supplementary Fig. 2). After engraftment, to evaluate the response of different numbers of MuSCs to injury, bioluminescence values for each mouse were normalized to the bioluminescence value obtained at the plateau of engraftment for that mouse. Mice were divided into two groups; one of these received NTX damage whereas the other did not. After NTX injury, transplanted cells underwent a second wave of ~80-fold expansion and, after a second injury with NTX, a third wave of ~100-fold expansion was observed, assessed as luciferase activity relative to the activity at the engraftment plateau. A peak was observed ~15 days post injury in each case (Fig. 3a, b and Supplementary Fig. 2), consistent with the timing of regeneration of 18-Gy-irradiated NOD/SCID muscles evident by morphology, by the presence of Myf5–β-gal$^+$ donor-derived cells and by the presence of regenerating myofibres expressing embryonic myosin heavy chain (Supplementary Fig. 3). This bioluminescence assay revealed a drop in cell number at the end of each regenerative wave of cell expansion, suggesting that cell death may counterbalance stem and progenitor cell proliferation to achieve homeostasis post injury. In contrast, luciferase activity in undamaged control mice remained

relatively constant after engraftment (Fig. 3a and Supplementary Fig. 2). These results demonstrate that transplanted MuSCs can respond to serial injury with successive waves of progenitor expansion. The magnitude of the response to two sequential notexin injections suggests that stem cell function persisted over time.

To determine the basis for the accumulation of the bioluminescence signal in the course of regeneration, we analysed cell proliferation and apoptosis. Transverse sections of transplanted tibialis anterior muscles at days 7, 13 and 19 days post injury, revealed that donor-derived cells were proliferating as shown by the expression of the proliferative marker Ki67 in GFP$^+$ donor-derived cells (Supplementary Fig. 4). Thus, the ongoing proliferation over the three-week time period post injury supports the progressive accumulation of new donor-derived cells over time detected by bioluminescence. In contrast with proliferation, apoptosis slowly increased, and was highest at 19 days, suggesting that cell death could account for the decline in bioluminescence signal (Supplementary Fig. 5).

A potential caveat in these experiments is that the increase in luciferase signal might be due not only to changes in cell number but also to changes in the expression of the luciferase gene or increased access of luciferin to luciferase$^+$ cells in a damaged tissue. To address these possibilities, muscles previously transplanted with myoblasts, as in Fig. 2e, were damaged by NTX injection. Instead of
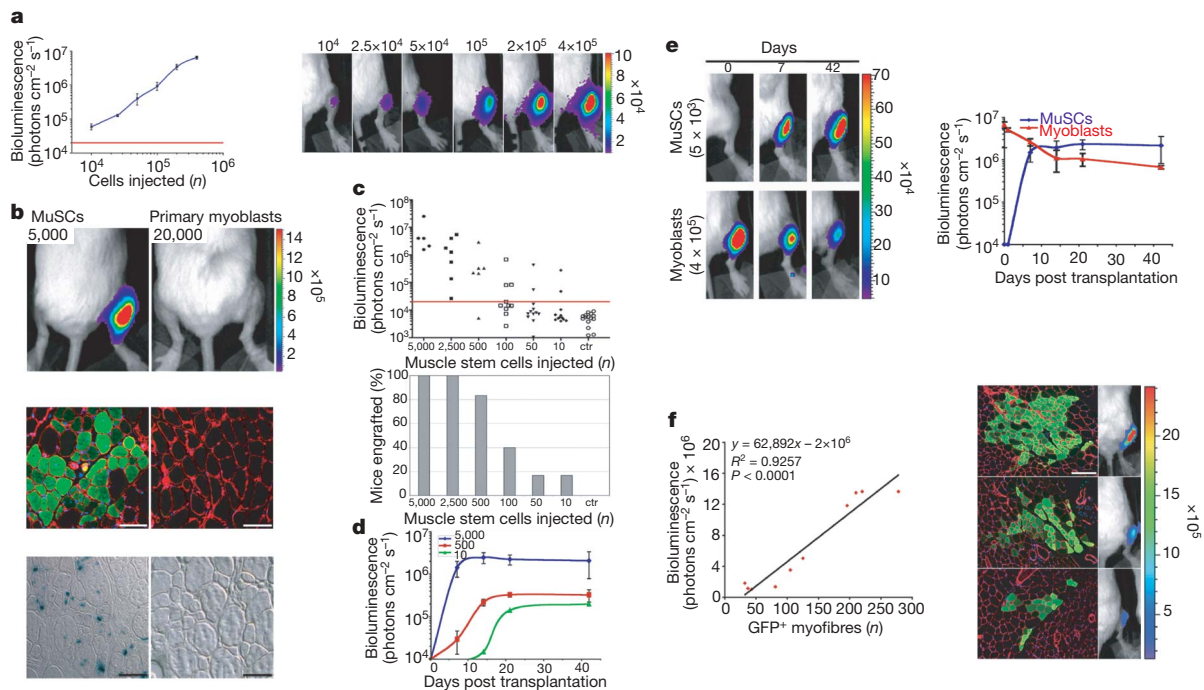

Figure 2 | MuSC engraftment monitored by *in vivo* non-invasive bioluminescence imaging. a, Increasing numbers of myoblasts isolated from *Myf5-nLacZ/FLuc* mice were injected into the tibialis anterior muscles of NOD/SCID recipients and imaged 2 h after injection. Shown is a graph of bioluminescence data represented as average ± s.e.m. ($n = 4$; $P \leq 0.05$; left) and a bioluminescent image of representative injected mice, with the number of cells injected indicated at the top (bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^4; right). **b**, Top, freshly isolated MuSCs or cultured myoblasts from double-transgenic mice were injected into recipients and imaged four weeks after transplantation ($n = 4$; $P < 0.005$; bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^5). Middle panels, immunofluorescence of luciferase expression in transverse muscle sections reveal the contribution of MuSCs, not of myoblasts, to muscle fibres. Laminin, red; luciferase, green; nuclei, blue. Scale bars, 100 μm. Bottom, 5 days before tissue harvesting, muscles were damaged with NTX. β-gal staining revealed Myf5–β-gal$^+$ cells in muscles transplanted with MuSCs, not with myoblasts. Scale bars, 100 μm. **c**, Different numbers of MuSCs were injected into muscles of recipients ($n = 50$), and engraftment was measured by imaging four weeks after transplantation (uninjected legs

are shown as negative control, ctr). A scattered graph of bioluminescent values of individual mice (top) and a histogram graph showing the percentages of mice with successful engraftment for each number of cells injected are shown (bottom). **d**, Engraftment of MuSCs (5,000, 500, 10) was monitored by imaging over a period of six weeks after transplantation (average ± s.e.m.; $n = 3$; $P < 0.05$). **e**, Freshly isolated MuSCs or cultured myoblasts were transplanted into recipients and engraftment was monitored by imaging over a period of six weeks after transplantation. Bioluminescent images of representative injected mice acquired at the indicated days are shown (bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^4; left). Graph of bioluminescence measurements (average ± s.e.m., $n = 5$; $P < 0.05$, right). **f**, MuSCs from *GFP/FLuc* transgenic mice were transplanted into recipients. Four weeks later, mice were analysed by bioluminescence imaging and for histology. Regression analysis shows a significant ($n = 10$, $P < 0.0001$) correlation between the number of GFP$^+$ myofibres and luciferase activity in individual mice (left). Representative images of immunofluorescence (GFP, green; laminin, red; nuclei, blue; scale bar, 120 μm) and bioluminescence imaging (bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^5; right).

the bioluminescence signal increasing, it remained constant (Fig. 3a), indicating that luciferase activity is not affected by tissue damage and confirming that myoblasts are unable to achieve detectable proliferation in response to injury (Fig. 3a). Finally, we compared *in vitro* undifferentiated myoblasts and tibialis anterior myofibres isolated from *FLuc* transgenic mice and found that, per microgram of DNA, enzyme activity did not differ (Supplementary Fig. 6 and Supplementary Methods). These results demonstrate that luciferase activity constitutes a readout of muscle cell numbers.

To establish definitively that the MuSCs are capable of self-renewal *in vivo*, transplantation and analysis of the progeny of a single cell are required, the gold standard established for haematopoietic stem cells[23,25]. Because more than one cell (10–500 cells per muscle) was transplanted in the experiments described above, we could not rule out the possibility that different MuSCs gave rise to the three successive waves of progenitor proliferation (during engraftment and after two NTX injections) without ever giving rise to another muscle stem cell. To test this possibility, we performed single MuSC transplantation experiments (Fig. 4a and Supplementary Methods). In each of 2 independent experiments, 3 out of a total of 72 single cells (4%) exhibited engraftment above background levels four weeks after transplantation (Fig. 4b). These results revealed that the

progeny of a single adult MuSC are capable of remarkable proliferation during engraftment, as in the six mice a signal approximately equivalent to 21,000–84,000 cells (~14–17 doublings) was detected. Notably, the bioluminescence assay enabled these single-cell studies, as detection of a 4% frequency using classical histological analyses would have been extremely labour-intensive. Histological analysis revealed that in those mice with a detectable bioluminescence signal, the progeny of a single cell differentiated and fused with mature myofibres (Fig. 4c). To determine whether self-renewal had occurred *in vivo*, we dissected the muscles and identified at least 50 donor-derived luciferase[+] cells per mouse expressing the satellite cell transcription factor Pax7 ($81 \pm 7\%$ of total re-isolated luciferase[+] cells) two months after transplantation of a single cell (Fig. 4d). These results demonstrate that a single integrin-α_7[+]CD34[+] MuSC can self-renew, giving rise to a population of mononucleated Pax7[+] cells that stably reside in recipient muscles.

Here we provide evidence that adult MuSCs have self-renewal capacity *in vivo* by demonstrating that after transplantation a single cell
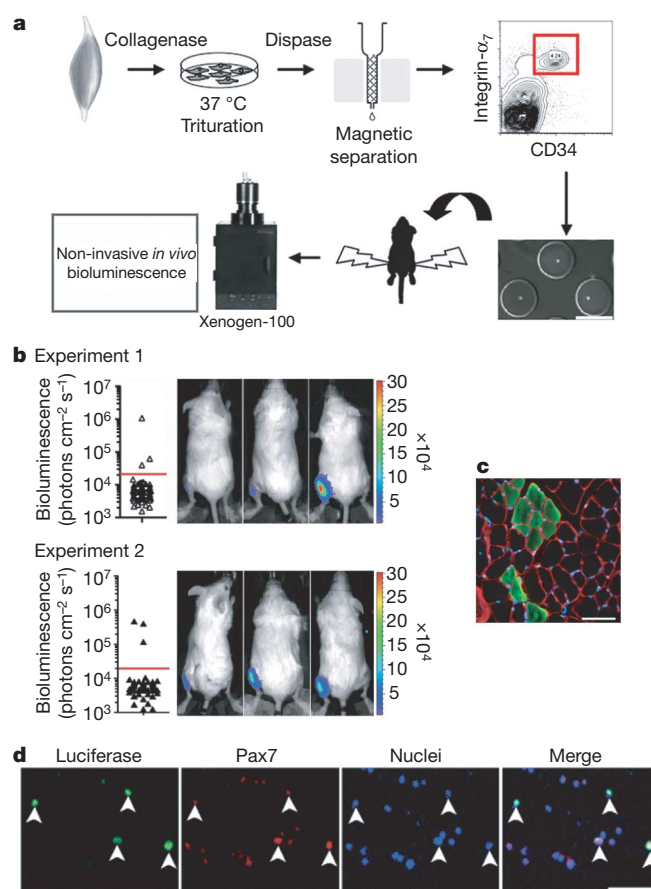

Figure 4 | Transplantation of a single MuSC demonstrates self-renewal capacity. a, Scheme of single MuSC transplantation. Cells were isolated from *Myf5-nLacZ/FLuc* transgenic mice by FACS as in Fig. 1a, segregated as single cells in hydrogel microwells (<10 μl per microwell; scale bars, 150 μm), individually picked by micromanipulation and transplanted into tibialis anterior muscles of recipient mice. **b**, Two independent experiments are shown ($n = 6/144$), and in each case 3 out of 72 single cell transplants resulted in engraftment above background detected by imaging mice four weeks after transplantation (left). Bioluminescent images of the 6 positive recipients after single MuSC transplantation (bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^4; right). **c**, Immunofluorescence of luciferase expression in transverse muscle sections from the mice in **b** revealed the contribution of single MuSC progeny to muscle fibres. Laminin, red; luciferase, green; nuclei, blue. Scale bars, 50 μm. **d**, Muscle cells were re-isolated from mice transplanted with single cells, and donor-derived luciferase[+]Pax7[+] cells were detected. Scale bars, 100 μm.

Figure 3 | MuSC proliferation in response to muscle tissue damage. a, A low number of MuSCs (10–500) was transplanted into recipients ($n = 10$). After the engraftment plateau was reached (as in Fig. 2d), mice were divided into two groups, and the tibialis anterior muscles of one group were damaged by NTX injection, resulting in a substantial increase in cell numbers. At 47 days after the first damage, a second NTX injection led to a second increase in cell number, whereas no significant change was detected in the undamaged group. As a comparison, high numbers of primary myoblasts (4×10^5) were transplanted into recipients and four weeks later muscles were damaged by NTX injection. The average bioluminescence fold increase ± s.e.m. is shown ($n = 5$; $P < 0.05$). **b**, Bioluminescent images of a representative animal transplanted with MuSCs and damaged with NTX acquired on the days indicated (top). Bioluminescence values are indicated as photons cm^{-2} s^{-1} ×10^4.

is capable of both producing copies of itself and generating more specialized progenitors. Moreover, the bioluminescence imaging used in this report revealed the time course and unexpected magnitude of the proliferative response of which adult MuSCs and their progeny are capable *in vivo*, in contrast to myoblasts (Supplementary Fig. 7). This assay should have broad application for quantitative comparisons among cell types after delivery to solid tissues in diverse disease and injury models.

METHODS SUMMARY

Muscle stem cell isolation. Tibialis anterior muscles of mice were subjected to enzymatic dissociation (first collagenase 0.2% and then dispase (0.04 units ml^{-1}, Sigma) for 90 min, after which non-muscle tissue was gently removed under a dissection microscope. The cell suspension was filtered through a $70\,\mu m$ nylon filter (Falcon) and incubated with the following biotinylated antibodies: CD45, CD11b, CD31 and Sca1 (BD Bioscience). Streptavidin beads (Miltenyi Biotech) were then added to the cells together with the following antibodies: integrin-α_7–phycoerythrin (PE; a gift from F. Rossi) and CD34–Alexa647 (eBioscience), after which magnetic depletion of biotin-positive cells was performed. The (CD45$^-$CD11b$^-$CD31$^-$Sca1$^-$)CD34$^+$integrin-α_7^+ population was then fractionated twice by flow cytometry (DIVA-Van, Becton-Dickinson). Primary myoblasts were isolated as described previously[20].

Bioluminescence imaging. For imaging, a Xenogen-100 device was used, as described previously[18]. In brief, the system comprises a light-tight imaging chamber, a charge-coupled device camera with a cryogenic refrigeration unit and the appropriate computer system (Living-Image Software; Xenogen). After intraperitoneal injection of luciferin in $100\,\mu l$ of PBS (0.1 mmol per kg body weight, Xenogen), we acquired images continuously for 30 min and stored them for subsequent analysis. We analysed images at 15 min after luciferin injection.

Immunofluorescence and histology. Muscle tissues were prepared for histology as described previously[17]. For immunofluorescence, rabbit anti-β-galactosidase (Molecular Probes), rabbit anti-luciferase (Abcam), rabbit anti-GFP (Molecular Probes), rat anti-laminin (Upstate Technologies), mouse anti-Pax7 (Developmental Mouse Hybridoma Bank), mouse anti-myogenin (Pharmingen), rat anti-Ki67 (DAKO), mouse anti-embryonic-myosin-heavy-chain (F1.652)[26] and TUNEL (ApopTAG Red kit, Chemicon) were used.

Statistical analysis. Data are presented as mean ± s.e.m. Comparisons between groups used the Student's *t*-test assuming two-tailed distributions, with an alpha level of 0.01–0.05.

1. Charge, S. B. & Rudnicki, M. A. Cellular and molecular regulation of muscle regeneration. *Physiol. Rev.* **84**, 209–238 (2004).
2. Mauro, A. Satellite cells of skeletal muscle fibers. *J. Biophys. Biochem. Cytol.* **9**, 493–495 (1961).
3. Montarras, D. *et al.* Direct isolation of satellite cells for skeletal muscle regeneration. *Science* **309**, 2064–2067 (2005).
4. Kuang, S. *et al.* Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**, 999–1010 (2007).
5. Cerletti, M. *et al.* Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134**, 37–47 (2008).
6. Collins, C. A. *et al.* Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell* **122**, 289–301 (2005).
7. Sherwood, R. I. *et al.* Isolation of adult mouse myogenic precursors: functional heterogeneity of cells within and engrafting skeletal muscle. *Cell* **119**, 543–554 (2004).
8. Tajbakhsh, S. *et al.* Gene targeting the *myf-5* locus with *nlacZ* reveals expression of this myogenic factor in mature skeletal muscle fibers as well as early embryonic muscle. *Dev. Dyn.* **206**, 291–300 (1996).
9. Seale, P. *et al.* Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**, 777–786 (2000).
10. Zammit, P. S. *et al.* Kinetics of myoblast proliferation show that resident satellite cells are competent to fully regenerate skeletal muscle fibers. *Exp. Cell Res.* **281**, 39–49 (2002).
11. Zammit, P. S. *et al.* Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? *J. Cell Biol.* **166**, 347–357 (2004).
12. Conboy, I. M. & Rando, T. A. The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. *Dev. Cell* **3**, 397–409 (2002).
13. Boutet, S. C. *et al.* Regulation of Pax3 by proteasomal degradation of monoubiquitinated protein in skeletal muscle progenitors. *Cell* **130**, 349–362 (2007).
14. Wakeford, S., Watt, D. J. & Partridge, T. A. X-irradiation improves mdx mouse muscle as a model of myofiber loss in DMD. *Muscle Nerve* **14**, 42–50 (1991).
15. Heslop, L., Morgan, J. E. & Partridge, T. A. Evidence for a myogenic stem cell that is exhausted in dystrophic muscle. *J. Cell Sci.* **113**, 2299–2308 (2000).
16. Harris, J. B. & Johnson, M. A. Further observations on the pathological responses of rat skeletal muscle to toxins isolated from the venom of the Australian tiger snake, *Notechis scutatus scutatus. Clin. Exp. Pharmacol. Physiol.* **5**, 587–600 (1978).
17. Sacco, A. *et al.* IGF-I increases bone marrow contribution to adult skeletal muscle and enhances the fusion of myelomonocytic precursors. *J. Cell Biol.* **171**, 483–492 (2005).
18. Wehrman, T. S. *et al.* Luminescent imaging of β-galactosidase activity in living subjects using sequential reporter-enzyme luminescence. *Nature Methods* **3**, 295–301 (2006).
19. Blau, H. M. & Hughes, S. M. Retroviral lineage markers for assessing myoblast fate *in vivo. Adv. Exp. Med. Biol.* **280**, 201–203 (1990).
20. Rando, T. A. & Blau, H. M. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J. Cell Biol.* **125**, 1275–1287 (1994).
21. Rando, T. A., Pavlath, G. K. & Blau, H. M. The fate of myoblasts following transplantation into mature muscle. *Exp. Cell Res.* **220**, 383–389 (1995).
22. Gussoni, E., Blau, H. M. & Kunkel, L. M. The fate of individual myoblasts after transplantation into muscles of DMD patients. *Nature Med.* **3**, 970–977 (1997).
23. Cao, Y. A. *et al.* Shifting foci of hematopoiesis during reconstitution from single stem cells. *Proc. Natl Acad. Sci. USA* **101**, 221–226 (2004).
24. Barberi, T. *et al.* Derivation of engraftable skeletal myoblasts from human embryonic stem cells. *Nature Med.* **13**, 642–648 (2007).
25. Osawa, M. *et al.* Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
26. Silberstein, L. *et al.* Developmental progression of myosin gene expression in cultured muscle cells. *Cell* **46**, 1075–1081 (1986).

Author Contributions A.S., H.M.B. and R.D. designed the research, and A.S. and R.D. performed the experiments with support from P.K. S.V. performed single-cell RT–PCR. A.S. analysed the data. A.S., R.D. and H.M.B. discussed the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.M.B. (hblau@stanford.edu).

nature

LETTERS

Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis

Djahida Bouskra[1], Christophe Brézillon[2], Marion Bérard[3], Catherine Werts[4,6], Rosa Varona[5], Ivo Gomperts Boneca[4,6] & Gérard Eberl[1]

Intestinal homeostasis is critical for efficient energy extraction from food and protection from pathogens. Its disruption can lead to an array of severe illnesses with major impacts on public health, such as inflammatory bowel disease characterized by self-destructive intestinal immunity. However, the mechanisms regulating the equilibrium between the large bacterial flora and the immune system remain unclear. Intestinal lymphoid tissues generate flora-reactive IgA-producing B cells, and include Peyer's patches and mesenteric lymph nodes, as well as numerous isolated lymphoid follicles (ILFs)[1,2]. Here we show that peptidoglycan from Gram-negative bacteria is necessary and sufficient to induce the genesis of ILFs in mice through recognition by the NOD1 (nucleotide-binding oligomerization domain containing 1) innate receptor in epithelial cells, and β-defensin 3- and CCL20-mediated signalling through the chemokine receptor CCR6. Maturation of ILFs into large B-cell clusters requires subsequent detection of bacteria by toll-like receptors. In the absence of ILFs, the composition of the intestinal bacterial community is profoundly altered. Our results demonstrate that intestinal bacterial commensals and the immune system communicate through an innate detection system to generate adaptive lymphoid tissues and maintain intestinal homeostasis.

Secondary lymphoid tissues, such as lymph nodes and Peyer's patches, are induced to develop in the sterile environment of the fetus by lymphoid tissue inducer (LTi) cells[3]. After birth, hundreds of small clusters of LTi-like cells termed cryptopatches (CPs) form between crypts in the intestinal lamina propria[4]. During bacterial colonization, several lines of evidence indicate that CPs recruit B cells and develop into ILFs[5,6], recapitulating the fetal development of lymph nodes and Peyer's patches[7]. However, in contrast to the programmed development of lymph nodes and Peyer's patches, the formation of mature ILFs from CPs requires induction by the gut flora[1,8]. Furthermore, in mice that harbour an expanded bacterial flora as a consequence of a lack of IgA, ILFs are abnormally large and numerous[9]. Given the inducible nature of ILFs, we have suggested that these lymphoid tissues play an essential role in the equilibrium between the gut flora and the immune system[7]. However, the components of the gut flora that induce the genesis of ILFs, the factors that induce the recruitment of B cells into CPs, and the impact of ILFs on the gut flora remain largely unknown.

Fetal LTi cells and LTi-like cells in CPs express the nuclear hormone receptor RORγt, which is required for the generation of LTi and CP cells and the development of secondary lymphoid tissues and ILFs[2,6,10]. The extensive network of CPs and ILFs was visualized in transgenic mice expressing the enhanced green fluorescent protein EGFP under control of the *Rorc(γt)* locus (*Rorc(γt)-Gfp*[TG] mice) on a bacterial artificial chromosome (BAC; Fig. 1a). CPs contain mostly RORγt[+] LTi cells and a smaller population of dendritic cells[6],

whereas ILFs constitute a continuum of structures harbouring increasing numbers of B cells (Fig. 1b)[5], previously grouped into immature ILFs (iILFs) that harbour relatively few B cells, and mature ILFs (mILFs) that form an organized B-cell follicle[11]. The number of CPs was highest in the proximal intestine, whereas ILFs, in particular mILFs, were most prominent in the distal ileum and the colon (Supplementary Fig. 1a), correlating with the increasing bacterial density towards the distal segments of the intestine. Furthermore, whereas CPs develop independently of germs[4] starting the second week after birth (Supplementary Fig. 1b), the development of ILFs required bacteria (Fig. 1c) and occurred around the weaning age of 2–3 weeks (Supplementary Fig. 1b), during a boost in bacterial input provided by solid food. In germ-free mice, only CPs or ILFs harbouring few B cells were found (data not shown)[1,8]. The development of CPs and the subsequent formation of ILFs was blocked by administration, during the first 2 weeks after birth, of lymphotoxin β receptor (LTβR)-Ig fusion protein (Supplementary Fig. 1c, d), an antagonist of lymphoid tissue development[12], demonstrating the developmental relationship between CPs and ILFs.

In order to identify which component of the intestinal flora induces the formation of ILFs, the flora of adult germ-free mice was first reconstituted with a cocktail of bacterial strains including *Escherichia coli* and members of the altered Schaedler's flora (ASF; Supplementary Fig. 2), derived from the normal commensal bacterial flora[13]. After 8 weeks, significant numbers of ILFs were found in the ileum and the colon (Fig. 1c), demonstrating that a restricted set of commensal bacteria species, devoid of invasive bacteria or pathogens, can induce the formation of ILFs. In these experimental conditions however, ileal ILFs failed to mature to the same extent as in specific pathogen-free (SPF) animals, suggesting that additional commensals are required for full maturation, or that the experimental colonization protocol used here in adult mice does not fully replicate natural colonization occurring in the ileum of newborn animals. We next assessed whether individual ASF species were sufficient to induce the formation of ILFs. Reconstitution of germ-free mice with the Gram[+] *Lactobacillus acidophilus* induced no or few ILFs, whereas reconstitution with the Gram[−] *Bacteroides distasonis* or *E. coli* induced high numbers of ileal and colonic ILFs (Fig. 1c). To confirm that Gram[−] commensals have a major impact on the formation of ILFs, SPF mice were treated from birth with antibiotics that target preferentially Gram[+] (vancomycin) or Gram[−] (colistin) bacteria (Fig. 1d). Similar to germ-free mice, mice treated with colistin formed few ILFs. In contrast, mice treated with vancomycin formed significant numbers of ILFs, but lower numbers than in control mice, suggesting some role for Gram[+] bacteria in the formation of ILFs, for example through activation of NOD2 and TLRs (see below), or some level of toxicity of vancomycin on Gram[−]

bacteria. Together, these results demonstrate a dominant role for commensal Gram⁻ bacteria in the formation of ILFs.

The requirement of bacteria in the formation of ILFs suggests the involvement of receptors of microbe-associated molecular patterns. We therefore investigated whether members of the TLR family and their associated signalling pathways, as well as members of the NOD-like
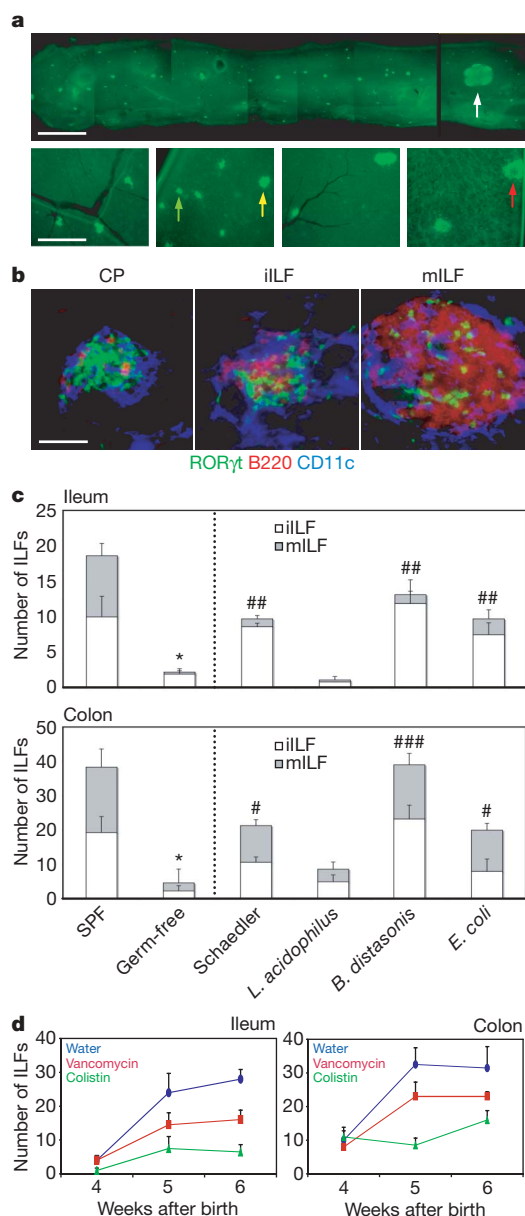
receptor family, were involved in the formation of ILFs. In the ileum and the colon of TLR2/4-, Myd88-, TRIF- or NOD2-deficient mice, maturation of ILFs was incomplete but induction of iILF formation was not significantly altered (Fig. 2a and Supplementary Fig. 3a). Thus TLRs and NOD2 appeared to be involved in the maturation of ILFs rather than in the induction of ILF formation. In accordance with this observation, TLRs activate the NF-κB pathway and induce the expression of inflammatory mediators such as TNF, required for the maturation of mILFs, but not for the induction of iILFs[11]. In contrast, the number of both iILFs and mILFs were significantly reduced in the ileum of NOD1-deficient mice (Fig. 2a), whereas CPs were hypertrophic (Fig. 2b). In the colon of NOD1-deficient mice, few iILFs were found whereas mILFs were hypertrophic (Supplementary Fig. 3a, b). Thus NOD1, recognizing peptidoglycans (PGNs) derived from Gram⁻ bacteria[14], was required for the generation of ileal iILFs, in accordance with our findings that Gram⁻ commensals were required in this process (Fig. 1c, d). Furthermore, using bone marrow chimaeras, we determined that NOD1 expression by stromal cells such as epithelial cells, but not haematopoietic cells such as dendritic cells and macrophages, were necessary for the generation of ileal ILFs (Fig. 2c). In the colon, the absence of NOD1 clearly affected ILF induction and maturation, but the large colonic flora appeared to induce ILFs also through pathways redundant with NOD1.

We next assessed whether a ligand for NOD1 derived from Gram⁻ bacteria, such as the muropeptide GMtetraDAP (N-acetylglucosamyl-β1,4-N-acetylmuramyl-L-alanyl-D-γ-glytamyl-meso-diaminopimelate-D-alanine) or its anhydro derivative also known as tracheal cytotoxin (TCT; Supplementary Fig. 4)[15], could induce the formation of ILFs. Germ-free mice were fed with GMtetraDAP or the NOD2 ligand

Figure 1 | Intestinal commensal Gram⁻ bacteria induce an extensive network of isolated lymphoid follicles (ILFs). a, The network of CPs and ILFs in the ileum of Rorc(γt)-Gfp[TG] mice. The upper right panel shows a Peyer's patch (white arrow) at the same magnification. The lower panels show, from left to right, proximal to distal regions of the small intestine. Arrows point to a CP (green), an immature ILF (iILF) (yellow) and a mature ILF (mILF) (red). Scale bars; 2.5 mm (upper panels), 500 μm (lower panels). **b**, A CP, a small iILF and a large mILF in sections from the ileum of a Rorc(γt)-Gfp[TG] mouse were stained as indicated. Scale bar, 60 μm. **c**, Quantification of ILFs in specific pathogen-free (SPF), germ-free mice, or mice reconstituted with the altered Schaedler's flora (ASF; Supplementary Fig. 2), L. acidophilus, B. distasonis or E. coli. At least 3 mice per group were analysed. Error bars, s.d.; *P < 0.001, unpaired t-test, total numbers of ILFs as compared with SPF mice; #P < 0.05, ##P < 0.01, ###P < 0.001, as compared with germ-free mice. **d**, ILFs in SPF mice of the indicated age that were treated from birth with the indicated antibiotics in drinking water. Error bars, s.d. At least 3 mice per group were examined.
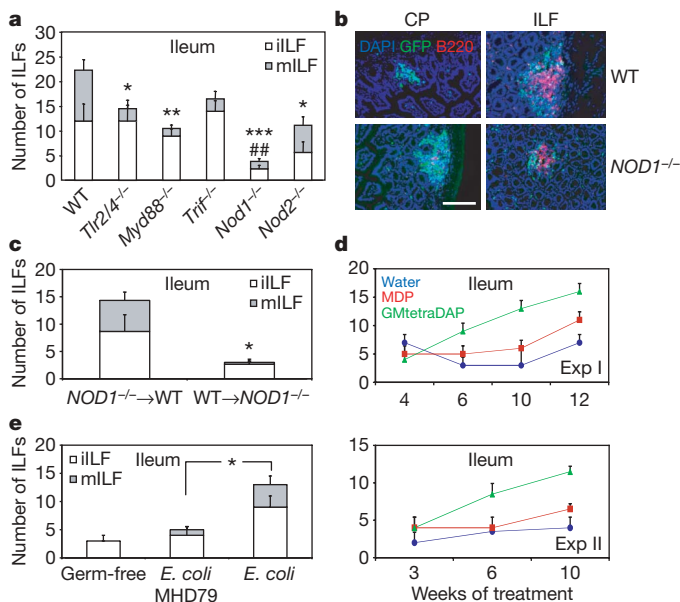
Figure 2 | Peptidoglycans recognized by NOD1 induce the formation of ILFs. a, The number of iILFs and mILFs in the distal ileum of wild-type (WT) or mutant mice. Error bars, s.d.; *P < 0.05, **P < 0.01, ***P < 0.001, unpaired t-test, total number of ILFs as compared with WT mice; ##P < 0.01, numbers of iILFs as compared with WT mice. **b**, CPs and ILFs in sections from the ileum of Nod1[+/−] × Rorc(γt)-Gfp[TG] (WT) mice or Nod1[−/−] × Rorc(γt)-Gfp[TG] mice were stained as indicated. Scale bar, 125 μm. **c**, Bone marrow cells from WT or Nod1[−/−] mice were adoptively transferred into the liver of irradiated newborn Nod1[−/−] or WT hosts, respectively. ILFs were counted 8 weeks after transfer. Error bars, s.d.; *P < 0.05, unpaired t-test, comparing total numbers of ILFs. **d**, Germ-free mice were given MDP or GMtetraDAP in the drinking water from 2 weeks of age (experiment I) or from birth (experiment II), and total ILFs were counted in the distal ileum after different periods of time. Error bars, s.d. **e**, Adult germ-free mice were reconstituted with E. coli MC1061 or the mutant E. coli MHD79, and ILFs counted 8 weeks later. Error bars, s.d.; *P < 0.05, unpaired t-test, comparing total numbers of ILFs. ILFs were counted in 2–4 mice per group.

muramyl-dipeptide (MDP) in their drinking water for different periods of time. GMtetraDAP, but not MDP, induced a significant number of ILFs, mostly iILFs, in the ileum after 6 weeks of treatment (Fig. 2d and Supplementary Fig. 5a). A significant number of ILFs was also detected in the colon after 10 weeks of treatment with GMtetraDAP starting from birth (Supplementary Fig. 3c, experiment II). Furthermore, when germ-free mice where reconstituted with an *E. coli* mutant (MHD79) that released low amounts of PGNs, containing mainly GMtetraDAP and TCT, as compared with wild-type *E. coli* MC1061, the induction of ILFs was abrogated (Fig. 2e and Supplementary Fig. 5b), demonstrating the essential role of NOD1 ligands in the induction of ILF formation by Gram⁻ bacterial commensals. Our finding that commensal bacteria, or TCT, induce the genesis of intestinal lymphoid tissues is reminiscent of the formation of the light-emitting organ in the squid *Euprymna scolopes*[16]. Colonization of internal crypts by the luminescent bacteria *Vibrio fischeri* induces haemocytes' (the invertebrate macrophages) infiltration of the crypts, epithelial apoptosis and morphogenesis, a process that is recapitulated by TCT alone. The squid molecule recognizing TCT remains however to be identified.

ILFs fail to form in mice deficient for the chemokine receptor CCR6 (Fig. 3a)[17], shown to be required on B cells and expressed by LTi-like cells in CPs[18]. CCR6 has two known ligands, CCL20 (or MIP-3α), which is highly expressed by the follicle-associated epithelium overlaying Peyer's patches[19] and ILFs (data not shown), as well as in ILFs[18], and β-defensin 3 (mBD3)[20], which is expressed by inflamed epithelia and intestinal crypts[21]. Interestingly, activation of NOD1 induces expression of hBD2, the human orthologue of mBD3, in a panel of human epithelial cell lines[22,23], and the expression of mBD4, another mouse orthologue of hBD2, is impaired in NOD1-deficient mice[24]. In mBD3-deficient mice, we found as few ILFs as in CCR6-deficient mice (Fig. 3a), demonstrating that mBD3 is required to induce the formation of ILFs, presumably through activation of its receptor CCR6. Furthermore, blocking CCL20 with neutralizing antibody also significantly reduced ILF formation, showing that the two CCR6 ligands have non-redundant roles in the activation of the CP-ILF system. Next, crypts were isolated from the ileum of germ-free or NOD1-deficient mice by laser capture microdissection (Supplementary Fig. 6), and transcripts (*Defb3*) coding for mBD3 or CCL20 were quantified by qRT–PCR. Whereas *Defb3* transcripts could not be detected by our method in crypts or in total intestinal epithelial cells, *Ccl20* transcripts were clearly detected and were substantially decreased in crypts isolated from germ-free or NOD1-deficient mice (Fig. 3b), indicating that commensals induced the expression of CCL20 through NOD1.
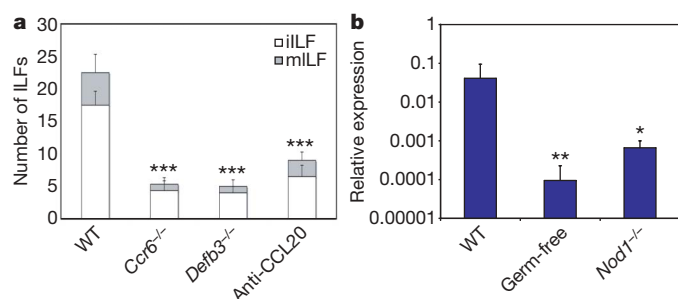
Together, these data suggest that the detection of bacterial commensals through NOD1 induces the expression of mBD3 and CCL20, which both activate CCR6 and induce the formation of iILFs. In patients, defective production of hBD2 predisposes to Crohn's disease[25], possibly both as direct and indirect consequences of the bactericidal and ILF-inducing activities of hBD2, respectively.

Finally, we measured the feedback impact of the CP-ILF system on the bacterial commensals. The bacterial biofilm of the ileum was collected from wild-type, NOD1- or mBD3-deficient mice, as well as from mice treated with LTβR-Ig fusion protein that lacked ILFs but developed a normal set of lymph nodes and Peyer's patches (Supplementary Fig. 1c). Quantitative PCR was performed on bacterial 16S rDNA using primers targeting the whole bacterial kingdom, or major groups of intestinal bacterial commensals (Supplementary Table 1). In both NOD1- and mBD3-deficient mice, the total bacterial population was expanded 100-fold, and in LTβR-Ig-treated mice, this expansion was tenfold (Fig. 4a, b). Specifically, the important groups of commensals *Clostridiales* and *Bacteroides*[9] and the smaller group of *Enterobacteriaceae* expanded 100-fold in NOD1-deficient mice and 10- to 100-fold in ILF-deficient mice, reminiscent of the results obtained in AID-deficient mice that do not produce IgA[9]. A small expansion of segmented filamentous bacteria (SFB), related to *Clostridiales* and potent inducers of Peyer's patch function[26], was observed in NOD1-deficient mice. In mBD3-deficient mice, the Gram⁻ *Bacteroides* and *Enterobacteriaceae* expanded, in accordance with the antimicrobial activity of the human orthologue hBD-2 against Gram⁻ bacteria[27]. On the other hand, the Gram⁺ *Lactobacillaceae* were reduced in all three types of mice, and *Clostridiales* failed to expand in mBD3-deficient mice that also lack ILFs, even though this group expanded in ILF-deficient mice. The individual ASF strains (Supplementary Fig. 2) followed, in general, the behaviour of their corresponding bacterial group (Supplementary Fig. 7). However, some exceptions underscore the complexity of behaviours of individual species or groups of bacteria in response to alterations both in the intestinal immune system and in the composition of the bacterial community. Together, these data show profound alterations in the bacterial flora of the ileum of mice that lack NOD1, mBD3 and/or ILFs, and demonstrate the role of the CP-ILF system in intestinal bacterial homeostasis. However, as expected, the absence of NOD1 or mBD3 had a stronger impact on
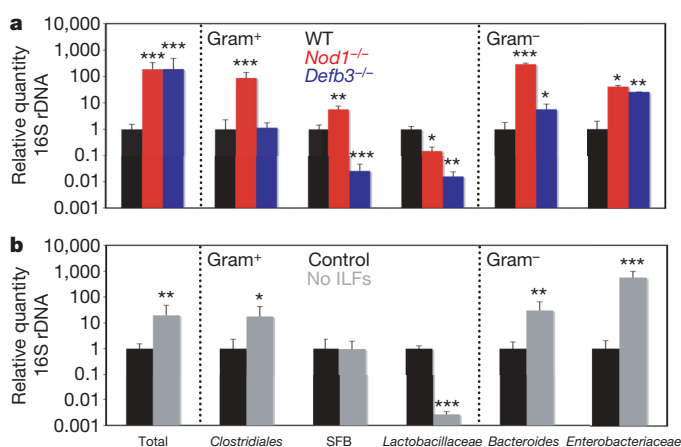
Figure 3 | A critical role for CCR6 ligands in the formation of ILFs. a, The number of iILFs and mILFs were counted in the distal ileum of *Ccr6*⁻/⁻ or *Defb3*⁻/⁻ (mBD3-deficient) mice, or mice injected with neutralizing anti-CCL20 antibody. ILFs were counted in 3–4 mice per group. **b,** Expression of transcripts coding for CCL20 in intestinal epithelium of WT, germ-free or *Nod1*⁻/⁻ mice. Total epithelium or the crypt regions were isolated by laser capture microdissection (Supplementary Fig. 6) and processed for RNA extraction. Results from duplicate PCR runs and two mice were normalized to the amplification of *Gadph* transcripts. Error bars, s.d.; *P < 0.05, **P < 0.01, ***P < 0.001, unpaired *t*-test, total numbers of ILFs or relative transcript expression, as compared with WT mice.

Figure 4 | The impact of the CP-ILF system on the commensal flora. a, b, Quantification by qPCR of total bacterial 16S rDNA, as well as of 16S rDNA from major groups of mouse intestinal commensals present in the ileal biofilm of WT, *Nod1*⁻/⁻ and *Defb3*⁻/⁻ mice (**a**) or of mice that lack ILF but have a normal set of lymph nodes and Peyer's patches (treated with LTβR-Ig fusion protein after birth), or in control mice (treated with human IgG1) (**b**). Data are normalized to values obtained in WT mice or control-treated mice. Error bars, s.d.; *P < 0.05, **P < 0.01, ***P < 0.001, unpaired *t*-test, as compared with WT or control-treated mice. 3–4 mice per group were analysed. All mice were bred in the same room of our mouse facility.

the total bacterial flora than the specific absence of ILFs in LTβR-Ig-treated mice, as in addition to decreased ILF formation, NOD1- (ref. 28) or mBD3-deficient mice have immune defects that directly impact on the bacterial flora, such as a decreased production of antimicrobial peptides in NOD1-deficient mice[24].

Our results demonstrate the reciprocal regulation of the intestinal bacterial flora and the ILFs, and the role of the CP-ILF system in intestinal homeostasis. Modulation of the CP-ILF system through the nuclear receptor RORγt expressed by CP cells, or administration of NOD1 ligands or Gram⁻ probiotics secreting high levels of NOD1 ligands, might therefore contribute to re-establishment of normal intestinal homeostasis during intestinal pathology. Interesting probiotic candidates include *E. coli* strain Nissle 1917, which strongly induces hBD2 in intestinal epithelial cells[29] and should therefore activate the CP-ILF system. Similarly, with the aim of using commensal bacteria or bacteria-derived compounds to treat intestinal disease, it has been shown[30] recently that a polysaccharide derived from the commensal *Bacteroides fragilis* effectively modulates intestinal inflammation. Interestingly, bacteria of the *Bacteroides* genus also induce the formation of ILFs (Fig. 1c).

METHODS SUMMARY

Mice. BAC-transgenic *Rorc(γt)-Gfp*[TG] mice were generated by inserting the coding sequence for EGFP, including the stop codon, into exon 1 of *Rorc(γt)* in place of the endogenous ATG translation start codon, on a 200 kb BAC (Invitrogen) carrying at least 70 kb of sequence upstream of the *Rorc(γt)* translation start site.

Quantification of CPs and ILFs. Two cm of distal ileum and 2 cm of proximal colon from 6–8 week old mice were flattened, fixed and frozen in OCT medium. 10–20 sections were cut at regular intervals through the whole thickness of the tissue and stained for CD45, CD11c and B220. CPs and ILFs were counted under a fluorescence microscope.

Peptidoglycan. PGN purified from *E. coli* strain MC1061 was digested with mutanolysin to generate the entire range of muropeptides present in *E. coli* PGN (Supplementary Fig. 3b). Soluble muropeptides were separated by reverse phase HPLC, and GMtetraDAP was collected and purity tested by a new HPLC analysis (Supplementary Fig. 3b). The nature of the GMtetraDAP was confirmed by MALDI-TOF mass spectrometry analysis.

Bacterial colonization of germ-free mice. Germ-free C3H/HeOrl mice were bred and housed in sterile isolators. Mice were inoculated at eight weeks of age by intra-gastric and intra-rectal administration of 10^7–10^8 colony forming units (c.f.u.) of ASF bacteria or *E. coli* as described by the Taconic Technical library (www.taconic.com/library).

Quantitative analysis of commensal bacteria. Quantitative PCR for 16S rDNA was performed on DNA purified from the ileal biofilm, and normalized to the total quantity of DNA recovered. In the case of ASF strains, absolute numbers of bacteria were determined from standard curves established by qPCR on serial dilutions of reference plasmids harbouring the 16S rDNA gene of each individual ASF strain.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hamada, H. *et al.* Identification of multiple isolated lymphoid follicles on the antimesenteric wall of the mouse small intestine. *J. Immunol.* **168**, 57–64 (2002).
2. Tsuji, M. *et al.* Requirement for lymphoid tissue-inducer cells in isolated follicle formation and T cell-independent immunoglobulin A generation in the gut. *Immunity* **29**, 261–271 (2008).
3. Mebius, R. E. Organogenesis of lymphoid tissues. *Nature Rev. Immunol.* **3**, 292–303 (2003).
4. Kanamori, Y. *et al.* Identification of novel lymphoid tissues in murine intestinal mucosa where clusters of c-kit⁺ IL-7R⁺ Thy1⁺ lympho-hemopoietic progenitors develop. *J. Exp. Med.* **184**, 1449–1459 (1996).
5. Pabst, O. *et al.* Cryptopatches and isolated lymphoid follicles: Dynamic lymphoid tissues dispensable for the generation of intraepithelial lymphocytes. *Eur. J. Immunol.* **35**, 98–107 (2005).
6. Eberl, G. & Littman, D. R. Thymic origin of intestinal αβ T cells revealed by fate mapping of RORγt⁺ cells. *Science* **305**, 248–251 (2004).
7. Eberl, G. Inducible lymphoid tissues in the adult gut: Recapitulation of a fetal developmental pathway? *Nature Rev. Immunol.* **5**, 413–420 (2005).
8. Pabst, O. *et al.* Adaptation of solitary intestinal lymphoid tissue in response to microbiota and chemokine receptor CCR7 signaling. *J. Immunol.* **177**, 6824–6832 (2006).
9. Fagarasan, S. *et al.* Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298**, 1424–1427 (2002).
10. Eberl, G. *et al.* An essential function for the nuclear receptor RORγt in the generation of fetal lymphoid tissue inducer cells. *Nature Immunol.* **5**, 64–73 (2004).
11. Lorenz, R. G., Chaplin, D. D., McDonald, K. G., McDonough, J. S. & Newberry, R. D. Isolated lymphoid follicle formation is inducible and dependent upon lymphotoxin-sufficient B lymphocytes, lymphotoxin β receptor, and TNF receptor I function. *J. Immunol.* **170**, 5475–5482 (2003).
12. Rennert, P. D., Browning, J. L., Mebius, R., Mackay, F. & Hochman, P. S. Surface lymphotoxin α/β complex is required for the development of peripheral lymphoid organs. *J. Exp. Med.* **184**, 1999–2006 (1996).
13. Dewhirst, F. E. *et al.* Phylogeny of the defined murine microbiota: Altered Schaedler flora. *Appl. Environ. Microbiol.* **65**, 3287–3292 (1999).
14. Fritz, J. H., Ferrero, R. L., Philpott, D. J. & Girardin, S. E. Nod-like proteins in immunity, inflammation and disease. *Nature Immunol.* **7**, 1250–1257 (2006).
15. Magalhaes, J. G. *et al.* Murine Nod1 but not its human orthologue mediates innate immune detection of tracheal cytotoxin. *EMBO Rep.* **6**, 1201–1207 (2005).
16. Koropatnick, T. A. *et al.* Microbial factor-mediated development in a host-bacterial mutualism. *Science* **306**, 1186–1188 (2004).
17. Lugering, A. & Kucharzik, T. Induction of intestinal lymphoid tissue: The role of cryptopatches. *Ann. NY Acad. Sci.* **1072**, 210–217 (2006).
18. McDonald, K. G. *et al.* CC chemokine receptor 6 expression by B lymphocytes is essential for the development of isolated lymphoid follicles. *Am. J. Pathol.* **170**, 1229–1240 (2007).
19. Tanaka, Y. *et al.* Selective expression of liver and activation-regulated chemokine (LARC) in intestinal epithelium in mice and humans. *Eur. J. Immunol.* **29**, 633–642 (1999).
20. Yang, D. *et al.* Beta-defensins: Linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* **286**, 525–528 (1999).
21. Bals, R. *et al.* Mouse beta-defensin 3 is an inducible antimicrobial peptide expressed in the epithelia of multiple organs. *Infect. Immun.* **67**, 3542–3547 (1999).
22. Voss, E. *et al.* NOD2/CARD15 mediates induction of the antimicrobial peptide human beta-defensin-2. *J. Biol. Chem.* **281**, 2005–2011 (2006).
23. Uehara, A., Fujimoto, Y., Fukase, K. & Takada, H. Various human epithelial cells express functional Toll-like receptors, NOD1 and NOD2 to produce anti-microbial peptides, but not proinflammatory cytokines. *Mol. Immunol.* **44**, 3100–3111 (2007).
24. Boughan, P. K. *et al.* Nucleotide-binding oligomerization domain-1 and epidermal growth factor receptor: Critical regulators of beta-defensins during *Helicobacter pylori* infection. *J. Biol. Chem.* **281**, 11637–11648 (2006).
25. Wehkamp, J. *et al.* Inducible and constitutive beta-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflamm. Bowel Dis.* **9**, 215–223 (2003).
26. Talham, G. L., Jiang, H. Q., Bos, N. A. & Cebra, J. J. Segmented filamentous bacteria are potent stimuli of a physiologically normal state of the murine gut mucosal immune system. *Infect. Immun.* **67**, 1992–2000 (1999).
27. Schroder, J. M. & Harder, J. Human β-defensin-2. *Int. J. Biochem. Cell Biol.* **31**, 645–651 (1999).
28. Fritz, J. H. *et al.* Nod1-mediated innate immune recognition of peptidoglycan contributes to the onset of adaptive immunity. *Immunity* **26**, 445–459 (2007).
29. Wehkamp, J. *et al.* NF-kappaB- and AP-1-mediated induction of human beta defensin-2 in intestinal epithelial cells by *Escherichia coli* Nissle 1917: A novel effect of a probiotic bacterium. *Infect. Immun.* **72**, 5750–5758 (2004).
30. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).

METHODS

Mice. BAC-transgenic $Rorc(\gamma t)$-Gfp^{TG} mice were generated as described previously[31]. The coding sequence for EGFP, including the stop codon, was inserted into exon 1 of $Rorc(\gamma t)$ in place of the endogenous ATG translation start codon, on a 200 kb BAC (Invitrogen) carrying at least 70 kb of sequence upstream of the $Rorc(\gamma t)$ translation start site. Germ-free mice of the C3H/HeOrl background (CDTA) were bred and maintained at the Institut Pasteur. MyD88-, TRIF-, TLR2- or TLR4-deficient mice were obtained from S. Akira and backcrossed at least for 8 generations into C57/Bl6J background. NOD1 ($card4^{-/-}$)-deficient mice were generated by Millennium Pharmaceuticals, and NOD2- ($card15^{-/-}$) deficient mice were provided by J.-P. Hugot. The $card15^{-/-}$ or $card4^{-/-}$ mice were backcrossed 8 generations into C57BL/6J mice[28]. All mice were kept in specific pathogen-free conditions and all animal experiments were approved by the committee on animal experimentation of the Institut Pasteur and by the French Ministry of Agriculture.

Antibodies, fusion proteins and antibiotics. Purified anti-CD11c (HL3) was purchased from BD Biosciences, phycoerythrin (PE)-, allophycocyanin (APC)- or biotin-conjugated anti-CD45R/B220 (RA3-6B2), and anti-CD45.2 (104) were purchased from eBiosciences, and purified anti-GFP (A-11122) and FITC-conjugated anti-rabbit polyclonal were purchased from Invitrogen. Neutralizing anti-CCL20 (114906) antibody was purchased from R&D Systems and 50–100 μg was injected i.p. in PBS twice a week. Human IgG1 and LTβR-Ig fusion protein have been previously described[12]; 100 μg protein was injected i.p. into pregnant mice 14.5 and 16.5 days after conception, or 50 μg protein was injected i.p. into newborn mice 1 day, 6 days and 12 days after birth, then 100 μg 15 days after birth. Vancomycin and colistin were purchased from Sigma and supplied to mice in the drinking water at a concentration of $500\,mg\,l^{-1}$ and $1\,g\,l^{-1}$, respectively.

Peptidoglycans. Peptidoglycans were purified from *E. coli* strain MC1061 as previously described[32]. Highly purified PGN (total of 100 mg) was digested with mutanolysin (Sigma)[33] to generate the entire range of muropeptides present in *E. coli* PGN (Supplementary Fig. 3b). Soluble muropeptides were separated by reverse phase HPLC, and the *N*-acetylglucosamyl-β1,4-*N*-acetylmuramyl-L-alanyl-D-γ-glytamyl-*meso*-diaminopimelate-D-alanine (GMtetraDAP) was collected and purity tested by a new HPLC analysis (Supplementary Fig. 3b). The nature of the GMtetraDAP was confirmed by MALDI-TOF mass spectrometry analysis as previously described[33]. Muramyldipeptide (MDP) was purchased from Sigma. Mice received MDP or GMtetraDAP in drinking water at a fixed concentration of 1 μM from the age of 2 weeks (experiment I) or from birth (experiment II) for different periods of time.

Bacterial strains and culture. ASF bacteria include the Gram$^+$ *Clostridium propionicum* (ASF 356), *Lactobacillus acidophilus* (ASF 360), *Lactobacillus murinus* (ASF 361), *Eubacterium plexicaudatum* (ASF 492), *Firmicutes* sp. (ASF 500) and *Clostridium* sp. (ASF 502), the Gram$^-$ *Mucispirillum schaedleri* (ASF 457) and *Bacteroides distasonis* (ASF 519) (Supplementary Fig. 2)[13,34], and were obtained from Taconic. The bacteria were cultured in anaerobic conditions on Schaedler agar supplemented with 5% sterile sheep blood (Biorad) in an anaerobic glove chamber containing a 5% CO_2, 10% H_2 and 85% N_2 atmosphere (Jacomex). The media were pre-reduced by placing them inside the chamber 2 days before inoculation of bacteria. Cultures were incubated in the chamber at 37 °C for different periods of time (1–5 days) depending on the bacterial strain. The hexa-mutant of *E. coli* lytic transglycosylase strain MHD79 was derived from the parental strain MC1061 and was previously described[35], and provided by J.-V. Höltje. Strains MC1061 and MHD79 are both streptomycin resistant (selection on LB agar with $100\,\mu g\,ml^{-1}$), and strain MHD79 carries additional resistance to chloramphenicol ($40\,\mu g\,ml^{-1}$), tetracycline ($12.5\,\mu g\,ml^{-1}$) and kanamycin ($50\,\mu g\,ml^{-1}$) that allows selective growth. *E. coli* cultures were obtained after 6 h at 37 °C in BHI broth (OD = 1 at 600 nm).

Bacterial colonization of germ-free mice. Germ-free C3H/HeOrl mice were bred and housed in sterile isolators. Mice were inoculated at eight weeks of age by intra-gastric and intra-rectal administration of 10^7–10^8 c.f.u. of ASF bacteria and *E. coli* as described by the Taconic Technical library (www.taconic.com/library). Faecal pellets from inoculated mice were assessed for colonization by microscopic examination of smears and quantified by serial dilutions of stool spread on Schaedler agar plates or by quantitative PCR (qPCR). Colonization of 8-week-old germ-free mice with wild-type *E. coli* (MC1061) or mutant *E. coli* (MHD79) was achieved by intra-gastric administration of 10^9 c.f.u. of bacteria. Bacterial status was checked every week. Fresh stool from each group were re-suspended in peptone broth and the suspensions were spread onto two series of LB agar plates supplemented with and without streptomycin ($100\,\mu g\,ml^{-1}$) to distinguish the *E. coli* strains from any other potential contaminant, or on LB agar plates supplemented with kanamycin ($50\,\mu g\,ml^{-1}$), a specific marker of strain MHD79. Mice were colonized to similar levels by strains MC1061 and MHD79 ($\sim 2.5 \times 10^9$ c.f.u. per g of faeces) and analysed 8 weeks after bacterial reconstitution.

Quantitative analysis of commensal bacteria. The whole ileum and colon were cut longitudinally. After removal of intestinal contents, the tissues were vigorously washed several times in PBS. The supernatants from these washes were used to quantify the bacterial content of the intestinal biofilm. DNA was extracted as described previously[36]. Quantitative PCR for 16S rDNA was performed on a Lightcycler apparatus (Roche) and normalized to total DNA. Absolute numbers of bacteria were determined from standard curves established by qPCR of serial dilutions of reference plasmids harbouring the 16S rDNA gene from each of the ASF strains. All reactions were performed in 10 μl reactions using the Quantitect SYBR green (Quiagen). Primers, reaction conditions and quantification have been adapted from three previous studies[37–39].

Generation of chimaeric mice. Bone marrow cells were isolated from C57BL/6 or $Nod1^{-/-}$ adult mice and treated with RBC buffer (Gibco) to eliminate red blood cells. A total of 3×10^6 cells in 25 μl of PBS were injected into the liver of C57BL/6 or $Nod1^{-/-}$ newborn mice that were previously irradiated at 400 rad. Reconstituted mice were killed and analysed 8 weeks after transfer.

Immunofluorescence histology. Intestines were opened and washed several times in PBS before fixation overnight at 4 °C in a fresh solution of 4% para-formaldehyde (Sigma) in PBS. The samples were then washed 1 day in PBS, incubated in a solution of 30% sucrose (Sigma) in PBS until the samples sank, embedded in OCT compound 4583 (Sakura Finetek), frozen in bath of iso-pentane cooled with liquid nitrogen and stored at −80 °C. Frozen blocks were cut at 8 μm thickness and sections collected onto Superfrost/Plus slides (VWR). Slides were dried for 1 h and processed for staining, or stored at −80 °C. For staining, slides were first hydrated in PBS-XG (PBS containing 0.1% triton X-100 and 1% normal bovine serum, VWR) for 5 min and blocked with 10% bovine serum in PBS-XG for 1 h at room temperature. Endogenous biotin was blocked with a biotin blocking kit (Vector Laboratories). Slides were then incubated with primary polyclonal Ab or conjugated mAb (in general 1/100) in PBS-XG overnight at 4 °C, washed 3 times for 5 min with PBS-XG, incubated with secondary conjugated polyclonal Ab or streptavidin for 1 h at room temperature, washed once, incubated with 4′6-diamidino-2-phenylindole-2HCl (DAPI) (Sigma) for 5 min at room temperature, washed 3 times for 5 min and mounted with Fluoromount-G (Southern Biotechnology Associates). Slides were examined under an AxioImager M1 fluorescence microscope (Zeiss) equipped with a CCD camera and images were processed with AxioVision software (Zeiss). For the quantification of CPs and ILFs, 10–20 sections were cut at regular intervals through the whole thickness of flat ileal or colonic fragments of 2 cm length and stained. CPs and ILFs were then counted under a fluorescence microscope.

Laser capture microdissection. Embryos or organs were embedded in OCT compound 4583 (Sakura Finetek), frozen in a bath of isopentane cooled with liquid nitrogen and stored at −80 °C. Frozen blocks were cut at 10 μm thickness and serial sections collected onto Superfrost/Plus slides (VWR). Sections were immediately fixed for 5 min in acetone at −20 °C, dried and stored at −80 °C, or thawed and immediately stained for 5 s with Histogen (Molecular Devices), washed briefly in RNase-free water supplemented with ProtectRNA (Sigma), dehydrated successively in one bath of 70% ethanol for 30 s, two baths of 95% ethanol for 1 min, two baths of water-free ethanol (VWR) for 2 min, and two baths of xylene for 5 min, and air-dried. Slides were transferred immediately into a Veritas Laser Capture Microdissector (Molecular Devices), and the regions of interest were microdissected and captured with Capsure Macro laser capture microdissection caps (Molecular Devices). RNA was isolated using the Picopure RNA Isolation kit (Molecular Devices), and its quality assessed using the 2100 Bioanalyser system (Agilent Technologies).

Gene expression analysis. 250–500 pg of high-quality total RNA was subjected to one linear mRNA amplification cycle using the MessageBooster Kit for qRT–PCR (Epicentre Biotechnologies). 50–100 ng of amplified mRNA was then converted to cDNA using Superscript III (Invitrogen). All procedures were performed according to the manufacturer's protocols. Real time PCR was performed using specific primers pairs (SuperArray Bioscience Corporation) on a PTC-200 thermocycler equipped with a Chromo4 detector (Bio-Rad Laboratories). Additional primer pairs specific for *Defb3* message were tested[40]. Data was analysed using Opticon Monitor software (Bio-Rad Laboratories).

31. Lochner, M. *et al. In vivo* equilibrium of proinflammatory IL-17+ and regulatory IL-10+ Foxp3+ RORγt+ T cells. *J. Exp. Med.* **205**, 1381–1393 (2008).
32. Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003).
33. Antignac, A. *et al.* Detailed structural analysis of the peptidoglycan of the human pathogen *Neisseria meningitidis. J. Biol. Chem.* **278**, 31521–31528 (2003).
34. Schaedler, R. W. & Dubos, R. J. The fecal flora of various strains of mice. Its bearing on their susceptibility to endotoxin. *J. Exp. Med.* **115**, 1149–1160 (1962).
35. Heidrich, C., Ursinus, A., Berger, J., Schwarz, H. & Holtje, J. V. Effects of multiple deletions of murein hydrolases on viability, septum cleavage, and sensitivity to large toxic molecules in *Escherichia coli. J. Bacteriol.* **184**, 6093–6099 (2002).

36. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63,** 2802–2813 (1997).

37. Barman, M. *et al.* Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infect. Immun.* **76,** 907–915 (2008).

38. Sarma-Rupavtarm, R. B., Ge, Z., Schauer, D. B., Fox, J. G. & Polz, M. F. Spatial distribution and stability of the eight microbial species of the altered Schaedler flora in the mouse gastrointestinal tract. *Appl. Environ. Microbiol.* **70,** 2791–2800 (2004).

39. Ge, Z. *et al.* Colonization dynamics of altered Schaedler flora is influenced by gender, aging, and *Helicobacter hepaticus* infection in the intestines of Swiss Webster mice. *Appl. Environ. Microbiol.* **72,** 5100–5103 (2006).

40. Rivas-Santiago, B. *et al.* β-Defensin gene expression during the course of experimental tuberculosis infection. *J. Infect. Dis.* **194,** 697–701 (2006).

LETTERS

Sarcolemma-localized nNOS is required to maintain activity after mild exercise

Yvonne M. Kobayashi[1,2,3,4], Erik P. Rader[1,2,3,4], Robert W. Crawford[1,2,3,4], Nikhil K. Iyengar[4], Daniel R. Thedens[5], John A. Faulkner[7], Swapnesh V. Parikh[4], Robert M. Weiss[4], Jeffrey S. Chamberlain[8], Steven A. Moore[6] & Kevin P. Campbell[1,2,3,4]

Many neuromuscular conditions are characterized by an exaggerated exercise-induced fatigue response that is disproportionate to activity level. This fatigue is not necessarily correlated with greater central or peripheral fatigue in patients[1], and some patients experience severe fatigue without any demonstrable somatic disease[2]. Except in myopathies that are due to specific metabolic defects, the mechanism underlying this type of fatigue remains unknown[2]. With no treatment available, this form of inactivity is a major determinant of disability[3]. Here we show, using mouse models, that this exaggerated fatigue response is distinct from a loss in specific force production by muscle, and that sarcolemma-localized signalling by neuronal nitric oxide synthase (nNOS) in skeletal muscle is required to maintain activity after mild exercise. We show that *nNOS*-null mice do not have muscle pathology and have no loss of muscle-specific force after exercise but do display this exaggerated fatigue response to mild exercise. In mouse models of nNOS mislocalization from the sarcolemma, prolonged inactivity was only relieved by pharmacologically enhancing the cGMP signal that results from muscle nNOS activation during the nitric oxide signalling response to mild exercise. Our findings suggest that the mechanism underlying the exaggerated fatigue response to mild exercise is a lack of contraction-induced signalling from sarcolemma-localized nNOS, which decreases cGMP-mediated vasomodulation in the vessels that supply active muscle after mild exercise. Sarcolemmal nNOS staining was decreased in patient biopsies from a large number of distinct myopathies, suggesting a common mechanism of fatigue. Our results suggest that patients with an exaggerated fatigue response to mild exercise would show clinical improvement in response to treatment strategies aimed at improving exercise-induced signalling.

To understand the molecular basis of the exercise-induced fatigue response, we studied genetically defined mouse models. We designed an integrative *in vivo* assay to test conscious mice, subjecting the mice to brief low-speed treadmill exercise followed by testing in an open-field activity chamber (see Methods). We first assessed two dystrophic mouse lines, *mdx* (model for Duchenne muscular dystrophy)[4] and *Sgca*-null (model for limb-girdle muscular dystrophy type 2D that is deficient for the gene encoding α-sarcoglycan (*Sgca*))[5]. In the absence of previous exercise, activity in these mice was indistinguishable from that of wild-type mice (Fig. 1a, b, and Supplementary Videos 1a–d). After a single trial of mild exercise, significant differences were observed (Fig. 1a, b, and Supplementary Videos 2a–d): the *mdx* and *Sgca*-null mice showed a significant decrease in vertical activity.

The decrease in vertical activity among *mdx* and *Sgca*-null mice did not correlate with differences in extensor digitorum longus (EDL)-specific force measurements relative to those taken in C57BL/6 mice before exercise (Fig. 1c). Moreover, *Sgca*-null mice do not develop brain, heart or vascular pathology[6], and they have muscle-force values similar to those of control mice[7]. Therefore, neither cardiac deficiency nor an inability to produce force was the cause of the post-exercise inactivity in the *Sgca*-null mice. Because inflammation is a feature of dystrophinopathy[4], chronic fatigue is associated with muscle pain, and chronic pain is associated with fatigue[8], we treated *mdx* mice with either deflazacort or ibuprofen. However, neither treatment resulted in improved post-exercise activity (Fig. 1d), suggesting that the inactivity occurring immediately after mild exercise in *mdx* mice was not due to inflammation or pain. Overall, the results of our exercise–activity assay implied that the exaggerated fatigue response in these mice was not attributable to cardiac deficiency, inflammation, pain or lack of muscle force.

To test whether the exercise-induced inactivity in the *mdx* and *Sgca*-null mice was due to the genetically determined structural defect in muscle, we assayed two mouse models in which the muscle pathology related to the specific dystrophin glycoprotein complex (DGC) defect is rescued—microdystrophin/*mdx* and MCKεSG/*Sgca*-null. In microdystrophin/*mdx* mice (a model for mild Becker muscular dystrophy[9]—the DGC has a mutated but functional dystrophin), microdystrophin is expressed in *mdx* mouse muscle. In the MCKεSG/*Sgca*-null mice, ε-sarcoglycan is expressed in mouse muscle that is deficient for *Sgca* (Supplementary Fig. 1). Neither rescue strain showed pathological signs of muscular dystrophy, and the skeletal muscle DGC of both was recovered at the biochemical, structural and functional levels (refs 9, 10 and Supplementary Figs 1 and 2).

Despite having a structurally intact skeletal muscle DGC, microdystrophin/*mdx* mice experience a substantial decrease in activity after mild exercise, like their *mdx* littermates (Fig. 1e). Because patients with Becker muscular dystrophy show profound fatigue after light exertion[11], and loss of sarcolemma-localized nNOS serves as a diagnostic indicator of some forms of Becker muscular dystrophy[12], a possible reason for the post-exercise inactivity is a loss of sarcolemma-localized nNOS. To test this possibility we probed for nNOS localization in microdystrophin/*mdx* skeletal muscle and found that the DGC generated in this rescue strain failed to recruit nNOS to the sarcolemma (Fig. 1e, inset). These data are in agreement with recent reports on microdystrophin expression in dystrophin-deficient mouse models[13]. Moreover, the data suggest that exercise-

[1]Howard Hughes Medical Institute, [2]Department of Molecular Physiology and Biophysics, [3]Department of Neurology, [4]Department of Internal Medicine, [5]Department of Radiology, [6]Department of Pathology, University of Iowa, Roy J. and Lucille A. Carver College of Medicine, 4283 Carver Biomedical Research Building, 285 Newton Road, Iowa City, Iowa 52242-1101, USA. [7]Department of Molecular and Integrative Physiology, University of Michigan, 2031 Biomedical Sciences Research Building, Ann Arbor, Michigan 48109-2200, USA. [8]Department of Neurology, University of Washington School of Medicine, HSB, Room K243b, Seattle, Washington 98195-7720, USA.
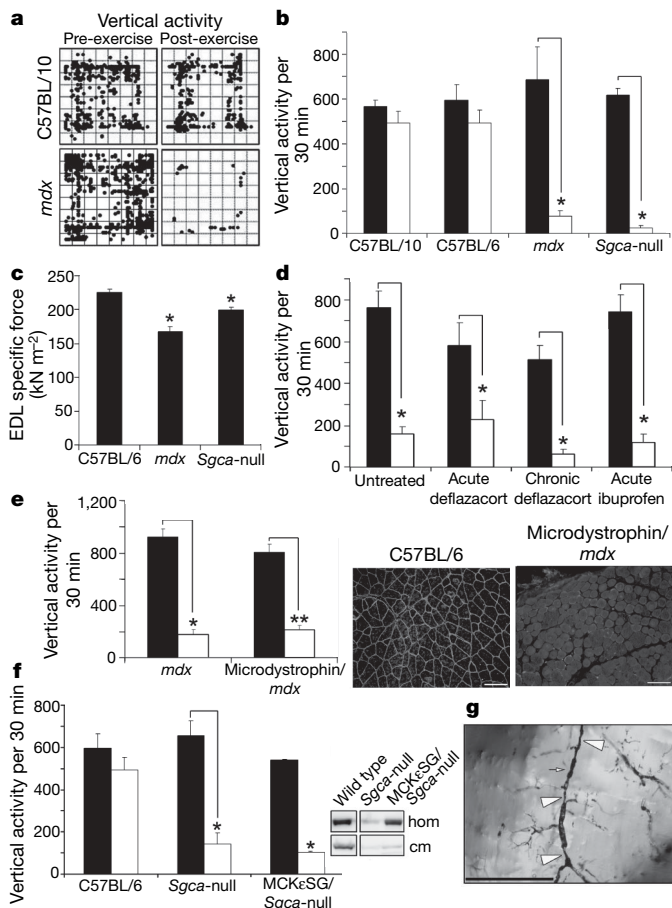
511

Figure 1 | Loss of sarcolemma-localized nNOS leads to skeletal muscle vascular narrowings, decreased capillary perfusion and an exaggerated fatigue response after mild exercise in dystrophic and non-dystrophic mouse models. a, Representative vertical activity tracings of zone maps for C57BL/10 and *mdx* mice before and after exercise. **b,** Quantified vertical activity before (filled columns) and after (open columns) exercise for C57BL/10, C57BL/6, *mdx* and *Sgca*-null mouse strains ($n = 6$ for each strain). Asterisk, $P = 0.012$. **c,** EDL muscle-specific force measurements from C57BL/6 ($n = 6$), *mdx* ($n = 4$) and *Sgca*-null ($n = 4$) mice. Asterisk, $P < 0.05$. **d,** Pre-exercise (filled columns) and post-exercise (open columns) vertical activity in untreated ($n = 7$) and anti-inflammatory treated *mdx* mice, acutely ($n = 4$) or chronically ($n = 4$) with deflazacort or acutely with ibuprofen ($n = 5$). Asterisk, $P < 0.003$. **e,** Left panel: quantified pre-exercise (filled columns) and post-exercise (open columns) vertical activity for microdystrophin/*mdx* mice ($n = 6$) and their *mdx* littermates ($n = 4$). Asterisk, $P = 0.005$; two asterisks, $P < 0.0001$. The right panels show representative immunofluorescence images of nNOS detection in the gastrocnemius muscles from C57BL/6 and microdystrophin/*mdx* mice. **f,** Quantified pre-exercise (filled columns) and post-exercise (open columns) vertical activity for MCKεSG/*Sgca*-null mice ($n = 6$) and their *Sgca*-null littermates ($n = 6$). Asterisk, $P < 0.0001$. Inset: immunoblot detection of total nNOS from homogenates (hom), and crude skeletal muscle membranes (cm). **g,** Representative Microfil image of skeletal muscle vessels of MCKεSG/*Sgca*-null mice after exercise—large arrowheads mark extended areas of vascular narrowing; the small arrow marks a shorter stretch of radial vascular narrowing. Error bars indicate s.e.m.

induced inactivity in the microdystrophin/*mdx* mice is not caused directly by a structurally defective muscle DGC, and that loss of sarcolemmal nNOS does not negatively affect muscle contractility. Thus, sarcolemmal nNOS seems to act at the level of post-exercise activity.

In contrast to the microdystrophin/*mdx* mice, MCKεSG/*Sgca*-null mice have structurally intact DGC in the brain and the vasculature, but express ε-sarcoglycan instead of α-sarcoglycan in the DGC of muscle. Our exercise–activity assay showed that post-exercise activity

in the MCKεSG/*Sgca*-null mice was substantially decreased relative to that in C57BL/6 mice but similar to that in *Sgca*-null and *mdx* mice (Fig. 1b, f). Because the microdystrophin-containing DGC failed to recruit nNOS, we speculated that the MCKεSG/*Sgca*-null mice would also fail to localize nNOS to the sarcolemma. Indeed, although total nNOS levels in muscle homogenates from MCKεSG/*Sgca*-null mice were similar to those in the wild type, nNOS from the rescue model failed to purify together with the ε-sarcoglycan-containing DGC in the membrane preparation (Fig. 1f, inset). Taken together, these results are compatible with the notion that the exaggerated fatigue response is not directly related to a structurally defective muscle DGC or to muscle weakness, but rather to a failure in the sarcolemmal localization of nNOS.

Because sarcolemma-localized nNOS is crucial for maintaining vasomodulation to contracting muscles[14], we tested whether communication from skeletal muscle to the local blood supply is deficient after mild exercise by perfusing MCKεSG/*Sgca*-null mouse arteries before or after exercise with Microfil and examined the skeletal muscle vasculature (Fig. 1g). We identified vascular narrowings of various lengths along the arteries that feed the skeletal muscles in the post-exercise samples only, and also noted the lack of perfusion of capillaries. The *mdx* and microdystrophin/*mdx* mice similarly showed skeletal muscle vascular narrowings only after exercise and also a lack of perfusion of capillaries (Supplementary Fig. 3c and data not shown). This phenotype is consistent with inefficient contraction-induced muscle nNOS signalling to local blood vessels. Overall, these data imply that loss of sarcolemma-localized nNOS causes deficient exercise-induced vasomodulation in skeletal muscle, and that these lead to prolonged inactivity after mild exercise.

To directly examine the contribution of NO generated by endothelial NOS (eNOS) or nNOS to the exaggerated fatigue response, we tested both *nNOS*-null and *eNOS*-null mice in our exercise–activity assay. Mice deficient for nNOS express normal levels of the DGC components at the sarcolemma and have histologically normal muscle[15–17]. Reports suggest that both mouse strains have defective vasoregulation[18,19]; however, *mdx* and *nNOS*-null mice have a normal α-adrenergic vasoconstrictive response to exercise[20]. Vertical pre-exercise activities were similar in *eNOS*-null, *nNOS*-null and C57BL/6 mice, suggesting that the loss of either NOS does not affect mouse activity (Fig. 2a). After exercise, however, *nNOS*-null vertical activity decreased significantly (Fig. 2a). Serum creatine kinase levels before and after exercise for each of the *NOS*-null mice were similar to those in C57BL/6 mice and low compared with *mdx* mice (Fig. 2b), and there were no signs of muscle pathology in sections from *nNOS*-null quadriceps muscle (Supplementary Fig. 4b), suggesting that muscle damage and necrosis were not the causes of the post-exercise inactivity. We then tested whether post-exercise muscle contractility affected the ability of C57BL/6 and *nNOS*-null skeletal muscle to produce force after mild exercise. We found that the specific force of EDL muscles after exercise was not significantly affected in *nNOS*-null muscle in comparison with C57BL/6 muscle (Supplementary Fig. 4c). Because lack of muscle contractility was not causing the inactivity in the *nNOS*-null mice after exercise, we checked whether *NOS*-null mice had post-exercise skeletal muscle vascular narrowings and lack of capillary perfusion similar to those in the dystrophic and rescue mice. Microfil perfusion of arteries from *NOS*-null mice before and after exercise revealed the lack of capillary perfusion and also the presence of vascular narrowings only in post-exercise *nNOS*-null skeletal muscle (Fig. 2c). We also found that treating wild-type mice with either the nNOS-specific inhibitor 3-bromo-7-nitroindazole or the vasoconstrictor sarafotoxin 6c caused post-exercise inactivity (Fig. 2d). These findings suggest that a deficiency of sarcolemma-localized nNOS causes exercise-induced narrowing of the vasculature that feeds active muscles after exercise, thereby promoting prolonged inactivity after mild exercise.

To test whether the vascular effect on post-exercise activity was from NO or was downstream of the NO signal, we bypassed

sarcolemmal nNOS signalling for decreasing vasoconstriction by treating *mdx* mice with a panel of pharmacological agents that promote vasodilation; we found that the exaggerated fatigue response was alleviated only by treatment with a phosphodiesterase (PDE) 5A inhibitor (Supplementary Fig. 6), suggesting that the fatigue that we saw depended on cGMP, which acts downstream of NO production. PDE activity in *mdx* mice is 2–6-fold higher than in C57BL/10 mice[21], which is consistent with the elevated PDE activity in human muscular disorders[18,21,22]. We treated *nNOS*-null, MCKεSG/*Sgca*-null and *mdx* mice with PDE5A inhibitors and tested them in our exercise–activity assay; we found that the treated MCKεSG/*Sgca*-null and *mdx* mice showed an increase in post-exercise activity (Fig. 2e and Supplementary Fig. 7a–d). Because inhibition of PDE5A had no effect on activity before exercise, our results suggest that PDE5A inhibition is alleviating the exaggerated fatigue response by enhancing the cGMP signal produced by contraction-induced nNOS stimulation. Although downstream effectors of cGMP are numerous

and divergent[23], the half-life of cGMP can be affected by the activity of PDE5A. Our data indicate that the elevated PDE activity in extracts from *mdx* mice could be PDE5A activity, and that PDE activity could also be elevated in the rescue mouse models we tested.

Our data suggest that the local resistance of arterioles that perfuse sarcolemmal nNOS-deficient muscles increases during exercise and that the lack of activity after mild exercise will lead to muscle oedema. We examined blood flow before and after exercise with laser Doppler imaging and found that blood flow in *mdx* mice failed to increase as it did in C57BL/6 mice (Fig. 3a and Supplementary Fig. 7a), but treatment of *mdx* mice with a PDE5A inhibitor alleviated this defect (Fig. 3b) and increased muscle capillary perfusion (Fig. 3c). Given that insufficient relief of local vasoconstriction in active muscles can lead to muscle oedema[24], and that boys with Duchenne muscular dystrophy show muscle oedema[25], we looked for changes in water compartmentalization and dynamics in the hindleg muscles of *nNOS*-null, C57BL/10 and *mdx* mice before and after exercise by using spin–spin relaxation time (T_2)-magnetic resonance imaging. The *nNOS*-null mice did not have muscle damage or loss of contractility after exercise (Fig. 2b and Supplementary Fig. 4b, c), nor did they have muscle oedema (Supplementary Fig. 8a), suggesting that their lack of muscle damage prevents water accumulation in the tissue. Similarly, C57BL/10 mice showed little to no oedema in hindleg muscles after exercise ($0.70 \pm 0.50\%$ (mean $\pm$ s.e.m.)) (Fig. 3f and
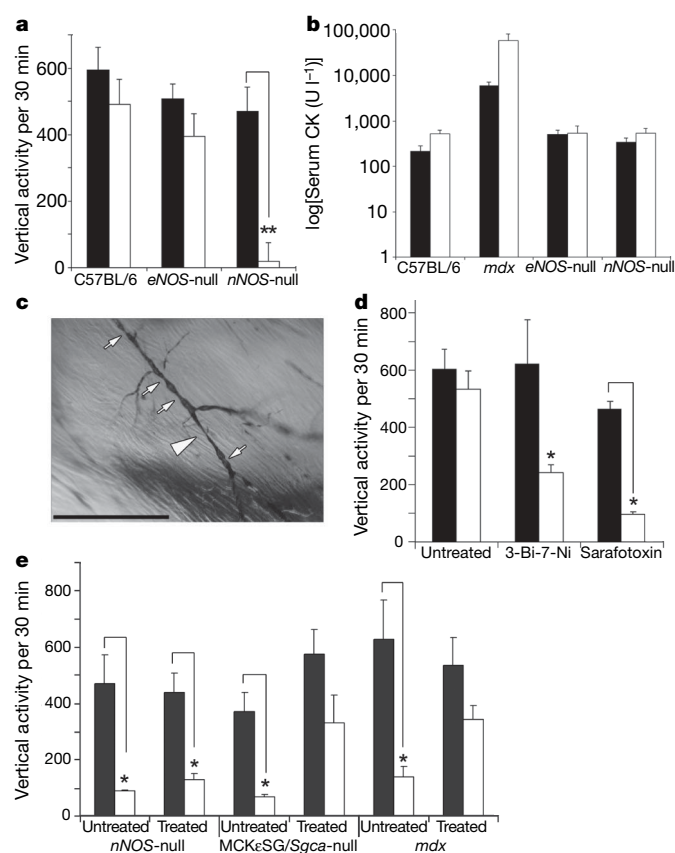
Figure 2 | Enhancing the cGMP signal resulting from muscle nNOS activation decreases the exaggerated fatigue response to mild exercise.
a, Comparison of vertical activity before (filled columns) and after (open columns) exercise between C57BL/6, *eNOS*-null and *nNOS*-null mice ($n = 6$ for each). Two asterisks, $P < 0.001$. **b,** Serum creatine kinase (CK) levels before (filled columns) and after (open columns) exercise in C57BL/6 ($n = 6$), *eNOS*-null ($n = 4$) and *nNOS*-null ($n = 6$) mice, compared with *mdx* mice ($n = 6$). **c,** Representative Microfil image of *nNOS*-null quadriceps skeletal muscle arteries after exercise—the large arrowhead marks the extended area of vascular narrowing; small arrows mark shorter areas of radial vascular narrowing. Scale bar, 100 μm. **d,** Pre-exercise (filled columns) and post-exercise (open columns) vertical activities in untreated wild-type (C57BL/6 and C57BL/10) mice ($n = 4$), compared with 3-B-7-Ni-treated wild-type mice ($n = 3$) and sarafotoxin-treated wild-type mice ($n = 4$). Asterisk, $P < 0.01$. **e,** Quantified pre-exercise (filled columns) and post-exercise (open columns) activity with and without treatment with PDE5A inhibitor, in *nNOS*-null ($n = 4$), MCKεSG/*Sgca*-null ($n = 4$) and *mdx* mice ($n = 6$). Asterisk, $P < 0.0001$. Pre-exercise and post-exercise vertical activity error bars are s.e.m.
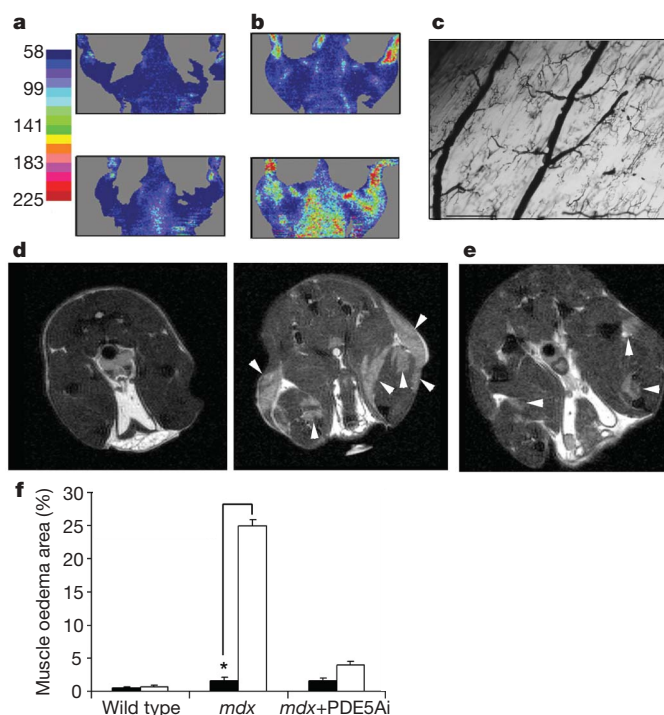
Figure 3 | Treatment with PDE5A inhibitor improves exercised-induced vasomodulation and decreases exercise-induced oedema in *mdx* mice.
a, Representative images of coronal laser Doppler analysis of blood flow in *mdx* mice before (top) and after (bottom) exercise ($n = 3$). **b,** Coronal laser Doppler analysis of blood flow in *mdx* mice, before (top) and after (bottom) exercise, treated with PDE5A inhibitor before exercise ($n = 3$). **c,** Representative Microfil image of quadriceps skeletal muscle arteries after exercise from *mdx* mice treated with PDE5A inhibitor before exercise ($n = 3$; scale bar, 100 μm). **d, e,** Representative axial views, by magnetic resonance imaging, of *mdx* hindlimb muscles before (left) and after (right) exercise ($n = 5$) (**d**) and hindlimb muscles after exercise of *mdx* mice treated with PDE5A inhibitor before exercise ($n = 5$) (**e**). White arrowheads mark areas of increased water compartmentalization. **f,** Percentage muscle oedema area before (filled columns) and after (open columns) exercise, and with or without treatment with PDE5A inhibitor, in *mdx* mice compared with that of the wild type. (Wild type and *mdx*, $n = 3$; *mdx* plus PDE5A inhibitor, $n = 5$; error bars are s.e.m.) Asterisk, $P < 0.001$.
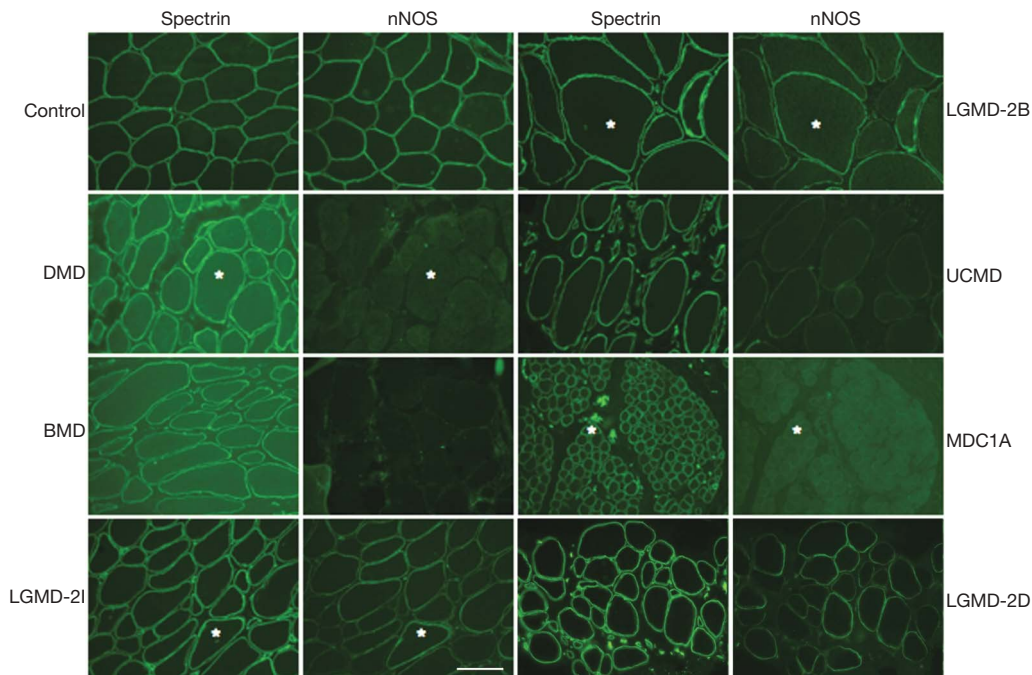
Figure 4 | nNOS levels in sarcolemma are decreased in human muscle diseases. Representative immunofluorescent staining in various human muscle diseases: primary dystrophinopathies, Duchenne and Becker muscular dystrophy (DMD and BMD, respectively); in several forms of limb-girdle muscular dystrophy (LGMD); in two congenital muscular dystrophies (CMD) caused by mutations in extracellular matrix proteins (Ullrich CMD (UCMD), collagen VI and merosin-deficient CMD (MDC1A). Asterisks mark the same muscle fibres in some of the adjacent panels. Scale bar, 100 μm.

Supplementary Fig. 8b). However, hindleg muscles of *mdx* mice consistently showed significant changes in tissue hydration after exercise ($25.0 \pm 2.45\%$ (mean $\pm$ s.e.m.)) (Fig. 3d, f) that were indicative of exercise-induced muscle oedema. The water accumulation observed in the *mdx* muscles is probably due to a combination of the increased local resistance in the arterioles that feed the active leg muscles and of muscle fibre fragility and damage. We also consistently found that treatment with PDE5A inhibitor significantly decreased exercise-induced muscle oedema in *mdx* mice ($3.99 \pm 0.82\%$ (mean $\pm$ s.e.m.); Fig. 3e, f). Overall, our data imply that treatment with PDE5A inhibitor can relieve the post-exercise inactivity by normalizing PDE activity, thereby allowing the available NO derived from muscle nNOS to signal for cGMP-dependent vasodilation in active muscle; treatment with PDE5A inhibitor decreases muscle damage in *mdx* mice by improving modulation of vascular activity in active muscle, thus preventing muscle oedema from exacerbating the muscle damage that occurs during the contraction of dystrophic muscle.

Because more than 60% of all patients with neuromuscular disease suffer from severe fatigue[2], we tested for nNOS localization to the sarcolemma in biopsies of patients representing different myopathic disorders (Fig. 4, Supplementary Fig. 9 and Supplementary Table 1). In most myopathic biopsies assessed, sarcolemma-localized nNOS was either reduced or not detected, implying that many myopathic disorders may share a mechanism that results in severe exercise-induced fatigue. Although increased fatigability inevitably occurs in patients with muscle weakness[1], our mouse data imply that the exercise-induced inactivity is distinct from muscle weakness and that loss of sarcolemma-localized nNOS leads to an exaggerated fatigue response to mild exercise.

Our mouse data show that decreased or mislocalized skeletal muscle nNOS exacerbates the fatigue experienced after mild exercise because the normal contraction-induced cGMP-dependent attenuation of local vasoconstriction fails to occur, and that this failure causes vascular narrowing in muscles after exercise. In addition, our data from *mdx* mice suggest that, as a result of nNOS mislocalization

and increased PDE activity[10,18,21], signalling for increased vasodilation to active muscle is deficient, causing muscle oedema. This, in turn, contributes to increased muscle damage as well as profound post-exercise debility. Although the exact mechanism that leads to the inactivity after mild exercise has not been reduced to a single beginning and end pathway, our data suggest that contraction-induced cGMP-dependent attenuation of local vasoconstriction is pivotal in this mechanism. These findings could lead to a better understanding of muscle fatigue under other physiological conditions in which muscle nNOS expression, localization or activity is affected.

METHODS SUMMARY

Mouse models. Animal care and procedures were approved and performed in accordance with the standards set forth by the National Institutes of Health and the University of Iowa Animal Care and Use Committee.

Treadmill exercise and activity monitoring. Animals were mildly exercised with an adjustable variable-speed belt treadmill from AccuPacer. Activity based on ambulatory behaviour was assessed in an open-field test.

1. Schillings, M. L. *et al.* Experienced and physiological fatigue in neuromuscular disorders. *Clin. Neurophysiol.* **118,** 292–300 (2007).
2. Zwarts, M. J., Bleijenberg, G. & van Engelen, B. G. Clinical neurophysiology of fatigue. *Clin. Neurophysiol.* doi:10.1016/j.clinph.2007.09.126 (2007).
3. Kalkman, J. S., Schillings, M. L., Zwarts, M. J., van Engelen, B. G. & Bleijenberg, G. The development of a model of fatigue in neuromuscular disorders: A longitudinal study. *J. Psychosom. Res.* **62,** 571–579 (2007).
4. Radley, H. G., De Luca, A., Lynch, G. S. & Grounds, M. D. Duchenne muscular dystrophy: focus on pharmaceutical and nutritional interventions. *Int. J. Biochem. Cell Biol.* **39,** 469–477 (2007).
5. Duclos, F. *et al.* Progressive muscular dystrophy in α-sarcoglycan-deficient mice. *J. Cell Biol.* **142,** 1461–1471 (1998).
6. Ozawa, E., Mizuno, Y., Hagiwara, Y., Sasaoka, T. & Yoshida, M. Molecular and cell biology of the sarcoglycan complex. *Muscle Nerve* **32,** 563–576 (2005).
7. Consolino, C. M. *et al.* Muscles of mice deficient in α-sarcoglycan maintain large masses and near control force values throughout the life span. *Physiol. Genomics* **22,** 244–256 (2005).
8. Yokoyama, T., Lisi, T. L., Moore, S. A. & Sluka, K. A. Muscle fatigue increases the probability of developing hyperalgesia in mice. *J. Pain* **8,** 692–699 (2007).

9. Harper, S. Q. *et al.* Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. *Nature Med.* **8**, 253–261 (2002).

10. Imamura, M., Mochizuki, Y., Engvall, E. & Takeda, S. I. ε-Sarcoglycan compensates for lack of α-sarcoglycan in a mouse model of limb-girdle muscular dystrophy. *Hum. Mol. Genet.* **14**, 775–783 (2005).

11. Phillips, B. A. & Mastaglia, F. L. Exercise therapy in patients with myopathy. *Curr. Opin. Neurol.* **13**, 547–552 (2000).

12. Torelli, S. *et al.* Absence of neuronal nitric oxide synthase (nNOS) as a pathological marker for the diagnosis of Becker muscular dystrophy with rod domain deletions. *Neuropathol. Appl. Neurobiol.* **30**, 540–545 (2004).

13. Judge, L. M., Haraguchiln, M. & Chamberlain, J. S. Dissecting the signaling and mechanical functions of the dystrophin–glycoprotein complex. *J. Cell Sci.* **119**, 1537–1546 (2006).

14. Thomas, G. D., Shaul, P. W., Yuhanna, I. S., Froehner, S. C. & Adams, M. E. Vasomodulation by skeletal muscle-derived nitric oxide requires α-syntrophin-mediated sarcolemmal localization of neuronal nitric oxide synthase. *Circ. Res.* **92**, 554–560 (2003).

15. Chao, D. S., Silvagno, F. & Bredt, D. S. Muscular dystrophy in *mdx* mice despite lack of neuronal nitric oxide synthase. *J. Neurochem.* **71**, 784–789 (1998).

16. Crosbie, R. H. *et al. mdx* muscle pathology is independent of nNOS perturbation. *Hum. Mol. Genet.* **7**, 823–829 (1998).

17. Suzuki, N. *et al.* NO production results in suspension-induced muscle atrophy through dislocation of neuronal NOS. *J. Clin. Invest.* **117**, 2468–2476 (2007).

18. Asai, A. *et al.* Primary role of functional ischemia, quantitative evidence for the two-hit mechanism, and phosphodiesterase-5 inhibitor therapy in mouse muscular dystrophy. *PLoS ONE* **2**, e806 (2007).

19. Huang, P. L. *et al.* Hypertension in mice lacking the gene for endothelial nitric oxide synthase. *Nature* **377**, 239–242 (1995).

20. Thomas, G. D. *et al.* Impaired metabolic modulation of α-adrenergic vasoconstriction in dystrophin-deficient skeletal muscle. *Proc. Natl Acad. Sci. USA* **95**, 15090–15095 (1998).

21. Bloom, T. J. Age-related alterations in cyclic nucleotide phosphodiesterase activity in dystrophic mouse leg muscle. *Can. J. Physiol. Pharmacol.* **83**, 1055–1060 (2005).

22. Bloom, T. J. Cyclic nucleotide phosphodiesterase isozymes expressed in mouse skeletal muscle. *Can. J. Physiol. Pharmacol.* **80**, 1132–1135 (2002).

23. Kass, D. A., Champion, H. C. & Beavo, J. A. Phosphodiesterase type 5: expanding roles in cardiovascular regulation. *Circ. Res.* **101**, 1084–1095 (2007).

24. Persson, J., Ekelund, U. & Grande, P. O. Endogenous nitric oxide reduces microvascular permeability and tissue oedema during exercise in cat skeletal muscle. *J. Vasc. Res.* **40**, 538–546 (2003).

25. Marden, F. A., Connolly, A. M., Siegel, M. J. & Rubin, D. A. Compositional analysis of muscle in boys with Duchenne muscular dystrophy using MR imaging. *Skeletal Radiol.* **34**, 140–148 (2005).

LETTERS

A fast, robust and tunable synthetic gene oscillator

Jesse Stricker[1]*, Scott Cookson[1]*, Matthew R. Bennett[1,2]*, William H. Mather[1], Lev S. Tsimring[2] & Jeff Hasty[1,2]

One defining goal of synthetic biology is the development of engineering-based approaches that enable the construction of gene-regulatory networks according to 'design specifications' generated from computational modelling[1–6]. This approach provides a systematic framework for exploring how a given regulatory network generates a particular phenotypic behaviour. Several fundamental gene circuits have been developed using this approach, including toggle switches[7] and oscillators[8–10], and these have been applied in new contexts such as triggered biofilm development[11] and cellular population control[12]. Here we describe an engineered genetic oscillator in *Escherichia coli* that is fast, robust and persistent, with tunable oscillatory periods as fast as 13 min. The oscillator was designed using a previously modelled network architecture comprising linked positive and negative feedback loops[1,13]. Using a microfluidic platform tailored for single-cell microscopy, we precisely control environmental conditions and monitor oscillations in individual cells through multiple cycles. Experiments reveal remarkable robustness and persistence of oscillations in the designed circuit; almost every cell exhibited large-amplitude fluorescence oscillations throughout observation runs. The oscillatory period can be tuned by altering inducer levels, temperature and the media source. Computational modelling demonstrates that the key design principle for constructing a robust oscillator is a time delay in the negative feedback loop, which can mechanistically arise from the cascade of cellular processes involved in forming a functional transcription factor. The positive feedback loop increases the robustness of the oscillations and allows for greater tunability. Examination of our refined model suggested the existence of a simplified oscillator design without positive feedback, and we construct an oscillator strain confirming this computational prediction.

The synthetic gene oscillator is based on a previously reported theoretical design[1] and was constructed using *E. coli* components (Fig. 1a). The hybrid promoter ($p_{lac/ara-1}$; ref. 14) is composed of the activation operator site from the *araBAD* promoter placed in its normal location relative to the transcription start site, and repression operator sites from the *lacZYA* promoter placed both upstream and immediately downstream of the transcription start site. It is activated by the AraC protein in the presence of arabinose and repressed by the LacI protein in the absence of isopropyl β-D-1-thiogalactopyranoside (IPTG). We placed the *araC*, *lacI* and *yemGFP* (monomeric yeast-enhanced green fluorescent protein) genes under the control of three identical copies of $p_{lac/ara-1}$ to form three co-regulated transcription modules (Supplementary Information). Hence, activation of the promoters by the addition of arabinose and IPTG to the medium results in transcription of each component of the circuit, and increased production of AraC in the presence of arabinose results in a positive feedback loop that increases promoter activity. However, the concurrent increase in production of LacI results in a linked negative feedback loop that

decreases promoter activity, and the differential activity of the two feedback loops can drive oscillatory behaviour[1,13].

The oscillator cells (denoted JS011) exhibited ubiquitous fluorescence oscillations over the entire run time of each experiment (at least 4 h). For example, at 0.7% arabinose and 2 mM IPTG, more
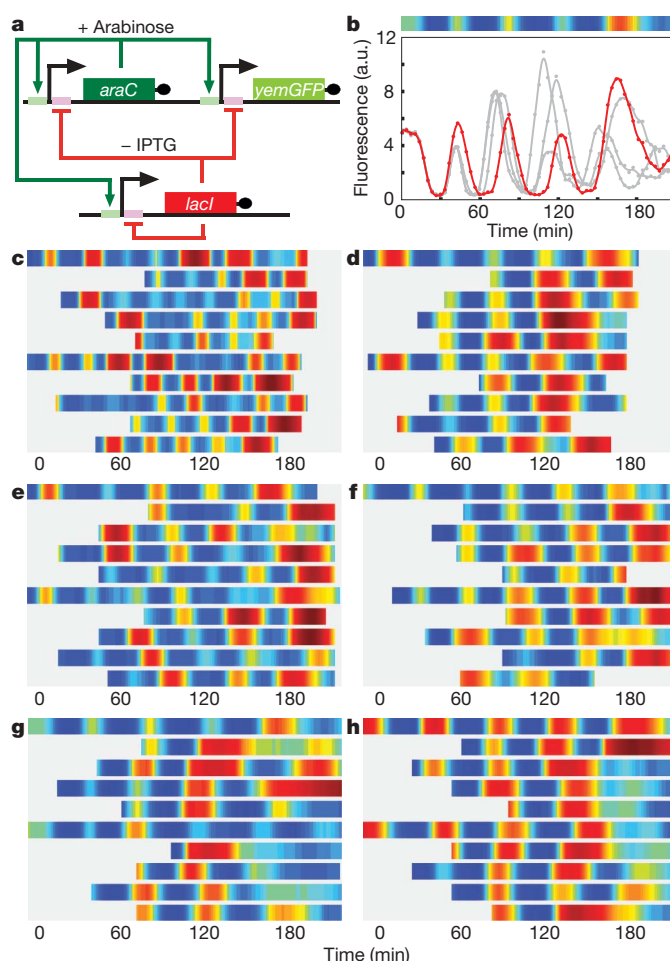

Figure 1 | Oscillations in the dual-feedback circuit. a, Network diagram of the dual-feedback oscillator. A hybrid promoter $p_{lac/ara-1}$ drives transcription of *araC* and *lacI*, forming positive and negative feedback loops. **b**, Single-cell fluorescence trajectories induced with 0.7% arabinose and 2 mM IPTG. Points represent experimental fluorescence values, and solid curves are smoothed by a Savitsky–Golay filter (for unsmoothed trajectories, see Supplementary Fig. 3). The trajectory in red corresponds to the density map above the graph. Density maps for trajectories in grey are shown in **g**. a.u., arbitrary units. **c–h**, Single-cell density map trajectories for various IPTG conditions (**c**, 0 mM IPTG; **d**, 0.25 mM; **e**, 0.5 mM; **f**, 1 mM; **g**, 2 mM; **h**, 5 mM).

[1]Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA. [2]Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

than 99% of the cells showed oscillations with a period of approximately 40 min (Fig. 1b, g, Supplementary Table 1 and Supplementary Movie 1). The highly dynamic nature of the oscillator components is shown by the rapid decay of green fluorescent protein (GFP) signal, which drops from peak to trough in less than 10 min (Fig. 1b). The oscillatory phase was heritable between daughter cells, which resulted in synchronized oscillations in areas of the microcolony derived from a common cell. This synchrony was limited to a few periods, presumably owing to oscillatory phase diffusion. We used a microfluidic device with a laminar boundary switch upstream of the growth chamber to investigate the initiation of synchronized oscillations (Supplementary Fig. 2c, d). Cells grown in the absence of inducer initiated oscillations in a synchronous manner on the addition of inducer (Supplementary Movie 10), which suggested the possibility of using flow cytometry to characterize the oscillator further. Flow cytometry of samples continuously collected from a culture in logarithmic growth that had been induced with 0.7% arabinose and 2 mM IPTG showed oscillations in mean cell fluorescence (Supplementary Fig. 8). Induction of oscillation was very quick (less than 5 min) and initially well-synchronized. The amplitude of these bulk oscillations decayed as the experiment progressed, as expected from the desynchronization of individual cells in the colony (Supplementary Information). However, the period obtained from the flow cytometry method (green data points in all figures) compared favourably to that obtained from single cells using microscopy (red data points in all figures).

The oscillator was extremely robust over an extensive range of inducer conditions and temperatures. At 0.7% arabinose and 37 °C, almost every observed cell oscillated (Supplementary Table 1) at all IPTG concentrations examined (Fig. 1b–h and Supplementary Movies 1–8). Varying the IPTG concentration allowed for the tuning of the oscillator period (Fig. 2a), particularly at low IPTG concentrations. The period decreased at high IPTG concentrations, and subsequent characterization of the promoter revealed that this non-monotonic behaviour is probably caused by IPTG interference with AraC activation[15] (Supplementary Information). The cell doubling time on the microfluidic device remained largely steady between experiments, ranging from 22.3 min to 27.6 min at 37 °C and showing little correlation to IPTG concentration ($R^2 = 0.132$). Individual cell fluorescence trajectories showed a gradual increase in oscillatory

period as the cells were imaged on the microfluidic device (Supplementary Fig. 4). This increase was not seen in doubling times, implying that the cells were not experiencing nutritional difficulties or environmental stress that might cause an alteration in oscillator behaviour.

To explore further the robustness of the oscillator, we investigated the effect of varying arabinose, temperature and the media source. At a fixed value of 2 mM IPTG and at 37 °C, the oscillatory period can be tuned from 13 min to 58 min by varying the arabinose level from 0.1% to 3.0% (Fig. 2b). Cells grown in the absence of arabinose did not express measurable levels of GFP in single-cell microscopy or flow cytometry experiments, and high levels of arabinose seemed to saturate the system. We observed sustained oscillations at a range of temperatures from 25 °C to 37 °C, with a decreasing period as a function of temperature (Fig. 2c). The cell doubling time also decreased with temperature, as expected, and the oscillatory period increased monotonically with cell doubling time (Fig. 2d). The oscillator also functioned in minimal A medium with $2 \, \mathrm{g \, l^{-1}}$ glucose (Fig. 2c, d). Although the cell doubling time in minimal medium was significantly longer than in LB-Miller formulation lysogeny broth (LB) (80–90 min versus 22–24 min at 37 °C), the period in the minimal medium was very similar to that in LB (Fig. 2c, d). This result, together with the strong dependence of the period on IPTG and arabinose concentration (at constant cellular doubling times), demonstrates that the synthetic oscillator is not strongly coupled to the cell cycle. The similar dependence of the period and the doubling time on the temperature seems to be due to the thermodynamic change of the rate constants affecting all cellular processes.

The oscillator was constructed according to design principles determined from previous theoretical work[1]. However, we found that this original model failed to describe two important aspects of the experiments. First, the model could not describe the observed functional dependence of the period on inducer levels. Second, and perhaps most importantly, because careful parameter tuning was necessary for oscillations in the original model, it was not able to describe the robust behaviour demonstrated in the experiments. This suggests that only a small region of inducer space should support oscillations, in contrast to the robust behaviour demonstrated in the experiments. These shortcomings forced a re-evaluation of the derivation of the oscillator equations, and led to a new computational model that more accurately described the experimental observations. The new model incorporates the same coupled positive and negative feedback architecture, but includes details that were omitted from the previous model. In particular, we found that directly modelling processes such as protein–DNA binding, multimerization, translation, DNA looping, enzymatic degradation and protein folding greatly increased the accuracy of the model. The result is a computational model that is very robust to parameter variations and correctly describes the dynamics of the oscillator for a large range of IPTG and arabinose concentrations (see Box 1 and Supplementary Information).

In examining our refined model, we discovered another region in parameter space that would support oscillatory behaviour. Our model predicted that a constantly activated system with repression controlled by a negative feedback loop could produce oscillations in the absence of positive feedback (Supplementary Fig. 19). It has been proposed that negative feedback gene networks can oscillate as long as there is delay in the feedback[16,17], and, although there is no explicit delay in our model, the intermediate steps of translation, protein folding and multimerization of LacI provide an effective form of delay[18] that is sufficient to support oscillations. We constructed this system (denoted JS013) in E. coli using a hybrid promoter, $p_{LlacO-1}$ (ref. 14), that is activated in the absence of LacI (or presence of IPTG) to drive both lacI and yemGFP expression (Fig. 3a). We observed oscillations in these cells when examined by single-cell microscopy under inducing conditions (Fig. 3b, Supplementary Fig. 5 and Supplementary Movie 11). These oscillations were not as distinct or regular as in the dual-feedback oscillator, and they did not always
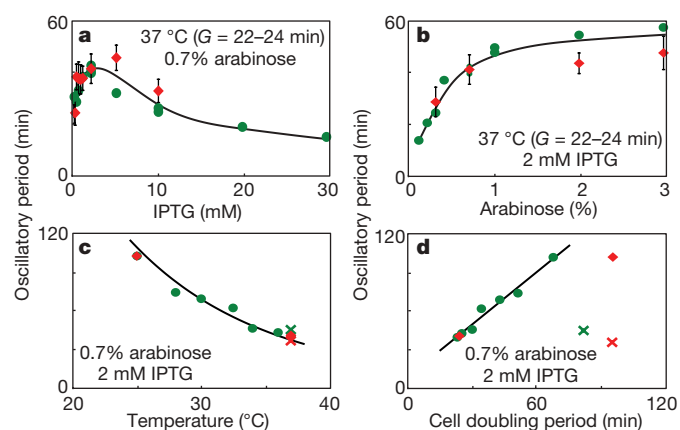

Figure 2 | Robust oscillations. a–c, Oscillatory periods on transects with 0.7% arabinose and varying IPTG (**a**), 2 mM IPTG and varying arabinose (**b**), or 0.7% arabinose, 2 mM IPTG, and varying temperature (**c**). Mean periods from single-cell microscopy (red diamonds, mean ± s.d.) or flow cytometry (green circles) are shown. Black curves are trend lines in **a** and **b**, or represent the theoretical prediction based on reference values at 30 °C in **c** (see Supplementary Information). Samples grown in minimal medium rather than LB are indicated by crosses. G represents the cell doubling period. **d,** Oscillatory period and cell division time increase monotonically as the growth temperature decreases. Symbols are as described above, and the black line is a linear regression of samples grown in LB.

Box 1 | Dynamic modelling of the dual-feedback oscillator circuit

We used standard techniques to construct both stochastic and deterministic computational models[3,25–28] based on the same underlying biochemical reactions illustrated in Fig. 4a (see Supplementary Information for full details of modelling). Although the interaction between transcription factors and the DNA is generally quite complicated to model in detail[29], we used experimental induction curves to calibrate the induction levels in the reactions describing the network (Supplementary Fig. 10). Over many oscillatory cycles, the deterministic simulations were then shown to give accurately the temporal evolution of the mode of the distributions generated by the exact stochastic simulations[24]. Representative time series for the protein concentrations obtained from the stochastic and deterministic models are depicted in Fig. 4b, c. The models are very robust in that oscillatory behaviour exists for a large range of parameter values and network details (Supplementary Information). Importantly, we found excellent quantitative agreement with the experimentally obtained period as a function of inducer levels (Fig. 4d, e).

The amplitude and period of the oscillations as a function of inducer levels can be conceptually explained using Fig. 4c. A burst begins with the basal transcription of messenger RNA from both promoters, encoding both the activator and the repressor. After a short delay (caused by, for example, translation, protein folding and multimerization), the amount of functional activator rises quicker than the amount of functional repressor, as shown in Fig. 4b. This occurs for two reasons. First, the activator gene is on a higher copy number plasmid than the repressor gene, meaning that more activator transcripts are produced than repressor transcripts. Second, assuming that transcription and translation of the monomeric forms of both proteins occur at similar rates, the activator will be more abundant because the functional tetrameric form of LacI requires twice as many monomers as does the functional dimeric form of AraC. As AraC levels rise, an activation burst in production of mRNA occurs due to the positive feedback loop. After LacI has been converted to a sufficient number of tetramers, the production of mRNA is turned off and the proteins decay enzymatically. Once all proteins have decayed, the promoters are freed of all bound regulators and the cycle begins anew. The length of the period is primarily determined by the time required for the proteins to decay. Therefore, the period is dependent on the rate of enzymatic decay and the magnitude of the activation burst. Furthermore, because the burst size depends on the induction characteristics of the promoter, it follows that the period is roughly proportional to the induction level of the promoter.

Figure 4 | Modelling the genetic oscillator. a, Intermediate processes are explicitly modelled in the refined oscillator model. **b, c**, Simulation results from Gillespie simulations (**b**) or deterministic modelling (**c**) at 0.7% arabinose and 2 mM IPTG. AraC dimers (green), LacI tetramers (red) and *lacI* mRNA (black) are shown. **d, e**, Comparison of modelling and experiment for oscillation period at 0.7% arabinose (**d**) or 2 mM IPTG (**e**). Values from deterministic modelling (blue curve), stochastic simulations (grey symbols, Supplementary Fig. 18), and microscopy (red diamonds) or flow cytometry (green circles) are shown. Lower and upper error bars in **d** represent the 16th and 84th percentiles, respectively, of the stochastic data, corresponding to ±1 s.d. for a normal distribution.

return to a dim state, consistent with the predictions of the computational model. Furthermore, the period was largely unaffected by IPTG concentration (varying less than 5% over three experimental runs from 0.6 mM to 20 mM IPTG), suggesting that the addition of the positive feedback loop serves the dual role of regularizing oscillations and allowing tunability of the period (see Supplementary Information).

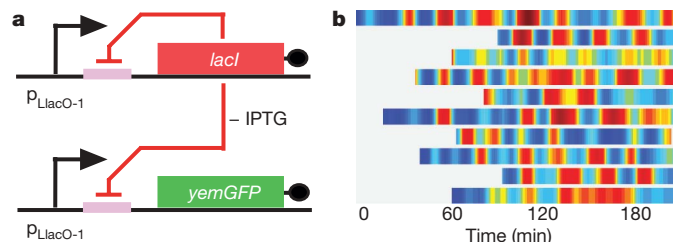In the context of synthetic biology, our findings indicate that caution must be exercised when making simplifying assumptions in the design of engineered gene circuits. We found that a full model of the system that takes into account intermediate steps such as multimerization, translation, protein folding and DNA looping is essential. The reason for this lies not only in the timescales of the system but also in the sequential timing of events. Because the intermediate steps in the production of functional protein take time, their introduction into the model creates an important form of delay[18–20]. We found that this effective delay greatly increases the robustness of our model. For instance, oscillatory activity in the model is only somewhat sensitive to the values chosen for system parameters (Supplementary Information), implying that nearly all cells should oscillate (Supplementary Table 1) despite minor stochastic variations in their intrinsic parameters. This determination of gene circuit design criteria in the present context of a fast, robust and tunable oscillator sets the stage for the design of applications such as expression schemes that are capable of circumventing cellular adaptability, centralized clocks that coordinate intracellular behaviour, and reverse-engineering platforms[21] that measure the global response of the genome to an oscillatory perturbation.

METHODS SUMMARY

The dual-feedback oscillator circuit was constructed by placing *araC*, *lacI* and *yemGFP* under the control of the hybrid p$_{lac/ara-1}$ promoter[14] in three separate transcriptional cassettes. An *ssrA* degradation tag[22] was added to each gene to decrease protein lifetime and to increase temporal resolution. These transcriptional cassettes were placed on two modular plasmids[14] and co-transformed into an ΔaraCΔlacI E. coli strain. The negative feedback oscillator circuit was constructed by placing *ssrA*-tagged *lacI* and *yemGFP* under the control of the p$_{LlacO-1}$ promoter[14] in two separate transcriptional cassettes, which were incorporated onto two modular plasmids and co-transformed into a ΔlacI strain. Cells were

Figure 3 | An oscillator with no positive feedback loop. a, Network diagram of the negative feedback oscillator. This oscillator is similar to the dual-feedback oscillator except that the hybrid promoter p$_{LlacO-1}$ (ref. 14) gives expression of *lacI* and *yemGFP* in the absence of LacI or in the presence of IPTG without requiring an activator. **b**, Single-cell density map trajectories for cells containing this oscillator (see Supplementary Movie 11 and Supplementary Fig. 5).
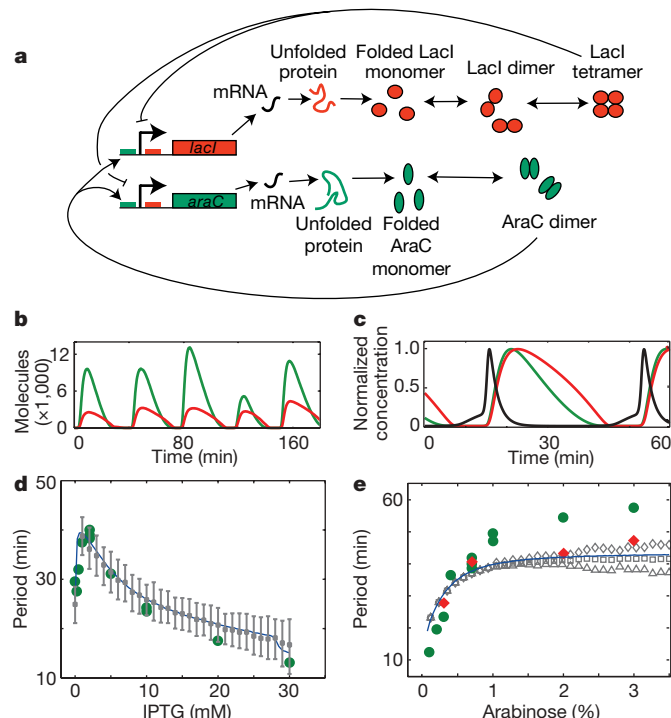
518

grown either in LB medium or in minimal A medium with $2\,g\,l^{-1}$ glucose. Oscillations were induced using arabinose (0.1–3%) and IPTG (0–30 mM). Single-cell microscopic data were collected by loading induced cells into polydimethylsiloxane-based microfluidic platforms that constrained the cells to a monolayer while supplying them with nutrients[23], and then providing a constant source of medium and inducers and imaging GFP fluorescence every 2–3 min for at least 4–6 h. These data were further analysed using ImageJ and custom-written Matlab scripts to extract single-cell fluorescence trajectories. Flow cytometry was performed either by taking samples from a continuously grown and serially diluted culture or by growing multiple cultures in parallel for varying durations. In either case, samples were read directly from their growth medium and low-scatter noise was removed by thresholding. Flow cytometry oscillatory periods were defined as the time elapsed between the first and second fluorescence peaks. Details of the models discussed are presented in Supplementary Information. Stochastic simulations were performed using Gillespie's algorithm[24], and deterministic simulations were performed using custom Matlab scripts.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hasty, J., Dolnik, M., Rottschäfer, V. & Collins, J. J. Synthetic gene network for entraining and amplifying cellular oscillations. *Phys. Rev. Lett.* **88**, 148101 (2002).
2. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
3. Tyson, J., Chen, K. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).
4. Sprinzak, D. & Elowitz, M. B. Reconstruction of genetic circuits. *Nature* **438**, 443–448 (2005).
5. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
6. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 2006.0028 (2006).
7. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
8. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
9. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
10. Fung, E. *et al.* A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).
11. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl Acad. Sci. USA* **101**, 8414–8419 (2004).
12. You, L., Cox, R. S., Weiss, R. & Arnold, F. H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
13. Barkai, N. & Leiber, S. Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
14. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
15. Lee, S. K. *et al.* Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl. Environ. Microbiol.* **73**, 5711–5715 (2007).
16. Bliss, R. D., Painter, P. R. & Marr, A. G. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* **97**, 177–193 (1982).
17. Bratsun, D., Volfson, D., Tsimring, L. S. & Hasty, J. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl Acad. Sci. USA* **102**, 14593–14598 (2005).
18. Rateitschak, K. & Wolkenhauer, O. Intracellular delay limits cyclic changes in gene expression. *Math. Biosci.* **205**, 163–179 (2007).
19. Mackey, M. & Glass, L. Oscillation and chaos in physiological control systems. *Science* **197**, 287–289 (1977).
20. Jaeger, J. & Reinitz, J. On the dynamic nature of positional information. *Bioessays* **28**, 1102–1111 (2006).
21. Faith, J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
22. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
23. Cookson, S., Ostroff, N., Pang, W. L., Volfson, D. & Hasty, J. Monitoring dynamics of single-cell gene expression over multiple cell cycles. *Mol. Syst. Biol.* **1**, 2005.0024 (2005).
24. Gillespie, D. T. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
25. Hasty, J. *et al.* Computational studies of gene regulatory networks: in numero molecular biology. *Nature Rev. Genet.* **2**, 268–279 (2001).
26. Ozbudak, E., Thattai, M., Lim, H., Shraiman, B. & van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli. Nature* **427**, 737–740 (2004).
27. Wang, X., Hao, N., Dohlman, H. & Elston, T. Bistability, stochasticity, and oscillations in the mitogen-activated protein kinase cascade. *Biophys. J.* **90**, 1961–1978 (2006).
28. Bennett, M. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–1122 (2008).
29. Gerland, U., Moroz, J. & Hwa, T. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl Acad. Sci. USA* **99**, 12015–12020 (2002).

Author Contributions J.S. and J.H. designed the oscillator circuits, and J.S. constructed the circuits. S.C. performed the microscopy experiments, and J.S. and S.C. performed the flow cytometry experiments. S.C., L.S.T. and J.H. performed the single-cell data analysis. M.R.B., W.H.M. and L.S.T. performed the computational modelling. All authors wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.H. (hasty@bioeng.ucsd.edu).

METHODS

Oscillator plasmid and strain construction. The oscillator components *araC* and *lacI* and a fluorescent reporter protein (*yemGFP*) were tagged with carboxy-terminal TSAANDENYALAA *ssrA* tags[22]. *yemGFP* contains F64L/S65T/A206K mutations. These tagged genes were then cloned into pZ modular plasmids under the transcriptional control of the $p_{lac/ara-1}$ hybrid promoter[14] to form three co-regulated transcriptional modules with identical promoters, ribosome-binding sequences and downstream terminators. The $p_{lac/ara-1}$ promoter is activated by AraC in the presence of arabinose and repressed by LacI in the absence of IPTG. The activator *araC* module and the reporter *yemGFP* module were placed on a ColE1 plasmid, and the repressor module was placed on a p15A plasmid. All PCR-amplified sections and sequence junctions were confirmed by sequencing. (See Supplementary Fig. 1.) An $\Delta araC \Delta lacI$ strain was constructed by P1*vir* phage transduction between $\Delta araC$ and $\Delta lacI$ strains. The two plasmids described above were co-transformed into this strain to construct the dual-feedback oscillator strain.

To construct the negative feedback oscillator strain, the hybrid promoter $p_{LlacO-1}$ (ref. 14) was used to regulate expression of *lacI* and *yemGFP*. This promoter is repressed by LacI in the absence of IPTG. Both genes were tagged with *ssrA* tags as described above. Two transcriptional modules containing $p_{LlacO-1}$ and *lacI* or *yemGFP* were constructed as above. The repressor module was placed on a p15A plasmid and the reporter module was placed on a ColE1 plasmid. These were then co-transformed into a $\Delta lacI$ strain.

Microscopy. We examined cells with single-cell time-lapse fluorescence microscopy using microfluidic devices designed to support growth of a monolayer of *E. coli* cells under constant nutrient flow (Supplementary Fig. 2). The design of the microfluidic device used in all microscopy experiments was adapted from the Tesla microchemostat design[23] for use with *Saccharomyces cerevisiae*. Modifications made to support imaging monolayers of *E. coli* included lowering the cell chamber height to match the cylindrical diameter of K-12 MG 1655 cells, lowering the delivery channel height to maintain equivalent flow splitting between the cell chamber and the bypass channel, and dividing the cell trapping region into three channels for simultaneous observation of isolated colonies (Supplementary Fig. 2a, b). For on-chip induction experiments, we used a variant of this device that incorporated a laminar boundary media switch into the design[30] and supported cell growth for several generations in non-inducing media before induction and imaging (Supplementary Fig. 2c, d).

In each experiment, a microfluidic device was mounted to the stage and wetted using a solution of 0.1% Tween 20 surfactant in the appropriate growth medium. For optimal *E. coli* growth, the chip temperature was typically maintained at 37 °C by flowing heated water through deep thermal channels fabricated into the device. Cells that had been passed from an overnight culture into inducing media approximately 3–4 h earlier were loaded into the device from the cell port by directing high flow both from the cell port and from the media port to the waste port. On trapping a single cell in each channel, flow past the cell chamber was reversed and slowed to 1–2 μm s^{-1} such that fresh nutrients were delivered from the media port by means of a combination of diffusion and advection without physically disturbing the cells.

Cells grew logarithmically to fill the channels over an experimental duration of ~4–6 h, while images were acquired every 2–3 min at ×100 magnification in the transmitted and fluorescent channels. Focus was maintained during image acquisition either by manual adjustment or by contrast-based autofocus algorithms. After each imaging session, fluorescence trajectories of individual cells were extracted using the WCIF ImageJ cell analysis package. For each fluorescence frame, mean values of integrated fluorescence were calculated within constant circular areas inscribed within the boundaries of all tracked cells. Long-term fluorescence trajectories were subsequently constructed by manually tracking each cell throughout the experiment.

Flow cytometry. Oscillator cells were initially characterized by flow cytometry of batch cultures to identify inducer conditions that supported oscillations. Subsequently, time-course flow cytometry was performed on growing cultures immediately after induction to follow oscillation dynamics. This time-course flow cytometry followed one of two similar protocols. In the continuous protocol, a single culture was serially diluted to maintain logarithmic growth. The culture was induced at the initial time point, and samples were removed for flow cytometry over the course of the experiment. In the aggregate protocol, an uninduced culture in logarithmic growth was aliquoted onto different inducer concentrations, and these subcultures were allowed to grow for varying lengths of time before flow cytometry. Flow cytometry was performed directly from growing cultures, and noncellular low-scatter noise was removed by thresholding. Oscillations were tracked by measuring the mean cellular fluorescence at each time point. The amplitude of the initial oscillation was usually higher than that of subsequent oscillations, presumably owing to desynchronization of the oscillations (Supplementary Fig. 8). The oscillation period was defined as the time elapsed between the first and second oscillation peaks. All flow cytometry analysis was carried out on a Becton-Dickinson FACScan.

30. Groisman, A. *et al.* A microfluidic chemostat for experiments with bacterial and yeast cells. *Nature Methods* **2**, 685–689 (2005).

LETTERS

Structural basis for gibberellin recognition by its receptor GID1

Asako Shimada[1]*, Miyako Ueguchi-Tanaka[1]*, Toru Nakatsu[2,3]*, Masatoshi Nakajima[4], Youichi Naoe[2],
Hiroko Ohmiya[1], Hiroaki Kato[2,3] & Makoto Matsuoka[1]

Gibberellins (GAs) are phytohormones essential for many developmental processes in plants[1]. A nuclear GA receptor, GIBBERELLIN INSENSITIVE DWARF1 (GID1), has a primary structure similar to that of the hormone-sensitive lipases (HSLs)[2,3]. Here we analyse the crystal structure of *Oryza sativa* GID1 (OsGID1) bound with GA$_4$ and GA$_3$ at 1.9 Å resolution. The overall structure of both complexes shows an α/β-hydrolase fold similar to that of HSLs except for an amino-terminal lid. The GA-binding pocket corresponds to the substrate-binding site of HSLs. On the basis of the OsGID1 structure, we mutagenized important residues for GA binding and examined their binding activities. Almost all of them showed very little or no activity, confirming that the residues revealed by structural analysis are important for GA binding. The replacement of Ile 133 with Leu or Val—residues corresponding to those of the lycophyte *Selaginella moellendorffii* GID1s—caused an increase in the binding affinity for GA$_{34}$, a 2β-hydroxylated GA$_4$. These observations indicate that GID1 originated from HSL and was further modified to have higher affinity and more strict selectivity for bioactive GAs by adapting the amino acids involved in GA binding in the course of plant evolution.

GAs comprise a large family of tetracyclic diterpenoid plant hormones that have diverse biological roles in plant growth. Although more than 130 GA derivatives have been identified, only a few GAs, such as GA$_4$, GA$_1$ and GA$_3$, actively function in plant growth[4]. Recent studies using GA-insensitive mutants of rice and *Arabidopsis* have revealed that three components—the GID1 nuclear receptor, DELLA repressor protein (SLR1 in rice) and F-box protein (GID2 in rice)—are important in GA perception[5,6]. Binding of GA to GID1 induces the formation of a GID1–GA–DELLA protein complex and, following the ubiquitin-ligase complex SCF^{GID2}-dependent degradation of DELLA protein, results in various GA-triggered actions. The binding affinity of GID1 to various GAs *in vitro* is well correlated with the effectiveness of each GA *in planta*, indicating that plant GA sensitivity essentially depends on the binding preference of GID1. The GID1–DELLA GA perception system is conserved among vascular plants but not in mosses[7,8]. However, GA specificity and sensitivity of the lycophyte *S. moellendorffii* GID1s (SmGID1s) are lower than those of GID1s in flowering plants, suggesting the optimization of the GA–GID1 interaction in the course of evolution of flowering plants.

The primary structural characteristic of GID1 is similar to that of the HSL family, which includes enzymes involved in lipid metabolism[9]. Proteins in the HSL family commonly contain an α/β-hydrolase fold of distinct topology. The catalytic centre contains a triad of amino acids (Ser-His-Asp). HSLs also have an additional mobile lid that controls access to the active site, triggering enhanced activity on contact with a lipid–water interface in the lipase reaction. Because the primary structures of GID1 and HSLs are similar, this raises the question of how different their tertiary structures might be and how GID1 manages to specifically interact with bioactive GAs while maintaining the conserved structure of the HSL family.

Here we report on the crystal structures of rice GID1 (OsGID1) interacting with GA$_4$ or GA$_3$. On the basis of the results of three-dimensional structure and interaction affinity analyses of different OsGID1s, each mutated at important residues for GA binding, we discuss how GID1 has acquired the GA reception ability that is lacking in HSLs.

We solved the crystal structures of GA$_4$–OsGID1 and GA$_3$–OsGID1 at 1.9 Å resolution. Because both structures were found to be essentially the same, here we focus on the GA$_4$–OsGID1 structure. As predicted by its sequence similarity to HSLs[3], GA$_4$–OsGID1 has an α/β-hydrolase fold similar to the catalytic domain of HSLs, namely, the central eight-stranded β-sheet (β1–β8) packed at both sides by helices (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 1). GA$_4$–OsGID1 also contains four helices—αb and αc at the N terminus and α4 and α5 at the central part between β6 and β7—that form lids. Two structural regions—the N-terminal 14 residues and the loop from Ala 85 to Thr 105 in between β2 and β3—are disordered and were not determined owing to weak electron density. A structural homology search by secondary-structure matching (SSM)[10] identified the closest homologue to be AeCXE1, a plant carboxylesterase from *Actinidia eriantha* (Protein Data Bank (PDB) entry, 2O7R)[11], which belongs to the HSLs (Fig. 1b).

The N-terminal lid consisting of αb and αc covers the central part of the carboxyl edge of the β-sheet (Fig. 1a and Supplementary Fig. 1). The cavity that extends under the lid of OsGID1 corresponds to the substrate-binding site of HSLs (Fig. 1b). GA$_4$ or GA$_3$ are held in their respective structures at the cavity above Ser 198, which is positioned at the apex of a sharp turn from β5 to α3 and is the nucleophilic elbow of the α/β-hydrolase fold[12] (Fig. 1c and Supplementary Fig. 2). This nucleophilic elbow is also present in HSLs, that is, Ser 169 at a turn from β5 to α4 in AeCXE1 (Fig. 1b and Supplementary Fig. 3). In this arrangement, the positive end of the helix dipole of α3 affects the polarizability of Ser 198. Thus, the side chain OH of Ser 198 probably forms strong hydrogen bonds with the carboxylate group at the C6 position in the B-ring of GA$_4$, indicating that this Ser is essential for GA binding. The corresponding Ser residue of HSLs is also essential as a nucleophile in the enzymatic function[13]. Aside from having this catalytic Ser, the residues corresponding to the catalytic triad of HSLs, namely, Ser, His and Asp, are similarly arranged except for the replacement of His with Val in OsGID1 (Supplementary Fig. 3). This replacement enables Val 326 to participate in GA binding as described later.
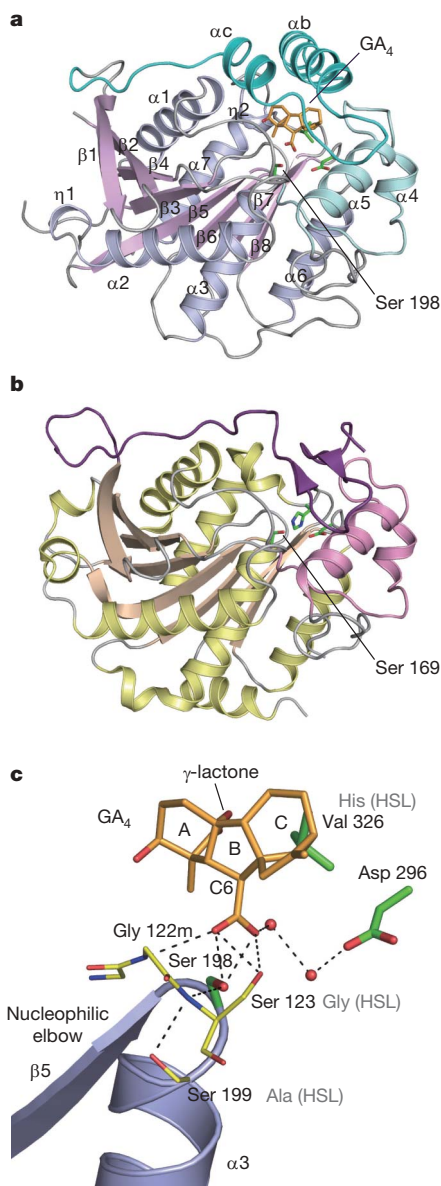
Figure 1 | Crystal structures of rice GID1 complexed with GA₄ and an HSL homologue. **a**, The GA₄–GID1 complex structure shown as a ribbon diagram. **b**, A ribbon diagram of plant carboxylesterase CXE1 from *A. eriantha* (PDB code: 2O7R). **c**, GID1 structure around the GA-binding site including corresponding residues for the HGGG (HGGS in GID1) motif (see Supplementary Discussion) and the catalytic triad of HSL (Ser 198, Asp 296 and Val 326). The residues of GID1 are indicated and the corresponding residues of AeCXE1 are also presented (HSL). The suffix 'm' represents that it is the main-chain atom. The rings of the *ent*-gibberellane structure are labelled as A, B and C, respectively, for GA₄. Water molecules are shown as red spheres.

The GA-binding pocket is formed by the N-terminal lid, mainly αb, and the scaffold that consists of the L120 loop (a loop comprising several amino acids around position 120 connecting β3 and α1), α1, the nucleophilic elbow, the carboxyl end of β6, the L240 loop (a loop connecting α4 and α5), α5, the L290 loop (a loop connecting β7 and α6), the L320 loop (a loop connecting β8 and η2) and η2 (Fig. 2a, b and Supplementary Figs 1 and 4). In this pocket, GA₄ is held directly by several hydrogen bonds with protein atoms or indirectly by means of a water molecule (Fig. 2a and Supplementary Fig. 4). The C6 carboxyl group of GA₄ forms the hydrogen bonds with the main-chain NH of Gly 122 and the side chains of Ser 198 and Ser 123. Ser 123 is stabilized by a network of hydrogen bonds formed with Arg 251 and Asp 250 in α5. The C6 carboxyl group also is linked to
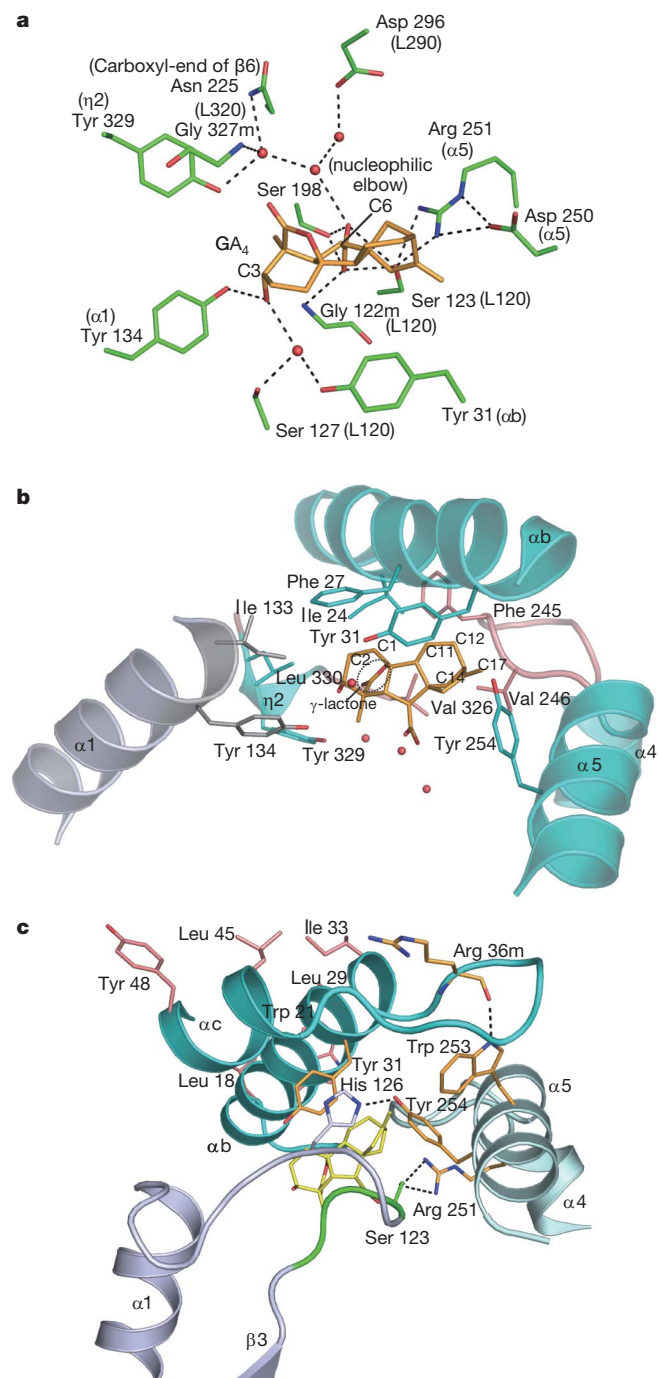

Figure 2 | Architecture of the GA-binding site in rice GID1. **a**, Polar residues that interact with GA₄. Water molecules are shown as red spheres. The positions in the secondary structural element are in parentheses. **b**, Residues with van der Waals or hydrophobic interaction with GA₄. **c**, Functional residues in the lid closure of GID1.

Tyr 329 in η2, Asn 225 at the carboxyl end of β6, and Asp 296 in L290 by means of two water molecules. The C3 hydroxyl group of GA₄ forms hydrogen bonds with Tyr 134 in α1, whereas it is linked indirectly to Ser 127 in L120 and Tyr 31 in αb by means of a water molecule. In the case of GA₃, there is an additional interaction between the C13 hydroxyl group and the Asp 250 in α5 by means of a hydrogen bond (Supplementary Fig. 2). No polar interaction was observed with the oxygen atoms of the γ-lactone ring of GA.

The molecular shape of the *ent*-gibberellane structure, the basal skeleton of all GAs, also contributes to keeping GA₄ held firmly in the pocket by non-polar interactions (that is, van der Waals and/or

hydrophobic interactions; Fig. 2b). Both faces of the gibberellane rings interact with Val 326 in L320 and Tyr 31 in αb. Val 326 is likely to recognize the surface structure of the γ-lactone ring. Tyr 31 comes into contact with the C17, C12 and C1 positions of GA$_4$ whereas Phe 27 in αb, Leu 330 in η2 and Ile 133 in α1 contact C1 and C2 in the A-ring. Tyr 254 in α5, Val 246 in L240 and Ile 24 in αb interact with C17, C14 and C11 in the C-ring, respectively. Because Ile 24, Phe 27 and Tyr 31 are the residues in the αb lid, the N-terminal lid holds GA$_4$ in the pocket, mostly by hydrophobic interactions. In other words, it is likely that GA$_4$ acts to close the N-terminal lid over the binding packet.

In the closed state of the N-terminal lid of the GA$_4$–OsGID1 complex, several interactions are observed between the lid and scaffold (Fig. 2c). Tyr 31 in αb and His 126 in L120 form an aromatic residue stack. His 126 is also stabilized by hydrogen bonding with Tyr 254 in α5. A loop between αb and αc in the lid captures Trp 253 in α5 by a van der Waals interaction and by a hydrogen bond between the main-chain CO of Arg 36 and the N atom in the indole ring of Trp 253. The interaction between Ser 123 in L120 and Arg 251 in α5 also confers stability to the closed state structure. The importance of the closed state of the N-terminal lid was confirmed by a marked decrease in the *in vitro* GA-binding activities of the mutated OsGID1s, H126A and W253A, in spite of no direct GA interaction of His 126 or Trp 253 (Supplementary Fig. 5). Several amino acid side chains in the N-terminal lid, such as Leu 18, Trp 21, Leu 29 and Ile 33 at αb, and Leu 45 and Tyr 48 at αc, extrude from the molecular surface (Fig. 2c). This region allows dimerization between two OsGID1 molecules in the crystal lattice (Supplementary Fig. 6), whereas OsGID1 behaves as a monomer during gel filtration (Supplementary Fig. 7). This led us to speculate that the hydrophobic exterior of the N-terminal lid might be involved in the interaction with SLR1. Replacement of the above six residues markedly diminished the interaction of OsGID1 with SLR1 in yeast cells (Supplementary Fig. 8), but was not accompanied by a decline in the GA-binding activity (Supplementary Fig. 5), supporting the above hypothesis.

To verify the structural assignment, we produced 17 OsGID1 mutants, each with different amino acid residues replaced with Ala, and examined their GA-binding activity *in vitro*. These mutants were newly produced and distinct from those analysed previously[3]. Thirteen mutated OsGID1s showed very little or no activity, whereas four retained only low or moderate activity (Fig. 3a and Supplementary Fig. 5). These observations confirm that these residues are important for GA-binding activity. Most of these residues are conserved within plant GID1s but not among HSLs (Supplementary Fig. 9), indicating that these residues have been recruited to establish a receptor for GA from the ancestral HSL structure. GID1 proteins in lycophyte *S. moellendorffii* (SmGID1s) sometimes contain non-conserved residues (Supplementary Fig. 9),
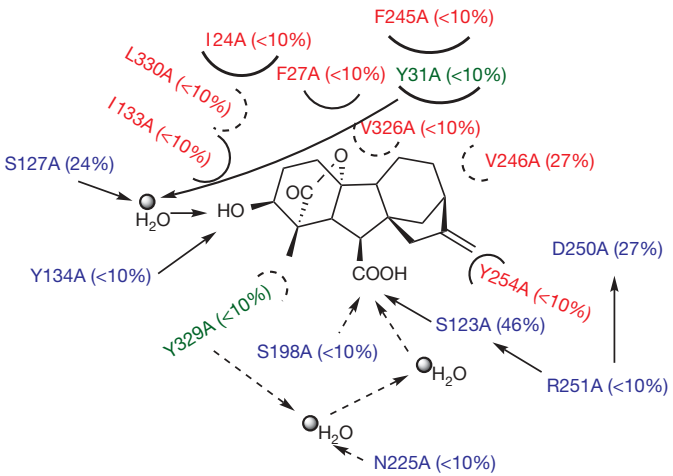

Figure 3 | GA binding activity of mutagenized rice GID1. The binding activities of mutants and their interaction sites with GA$_4$, relative to the native OsGID1 protein. Polar and non-polar interactions are indicated by arrows and circles, respectively. The interactions occurring behind GA are shown as dashed lines. Mutants with substitutions of the residues involved in polar, non-polar and both interactions are shown in blue, red and green, respectively.

showing lower affinity and specificity for GAs than GID1s in flowering plants[7]. We hypothesized that such amino acid replacements in SmGID1s may have led to their unique characteristics—that is, lower affinity and specificity for GAs. To examine this hypothesis, we exchanged the amino acid residues of OsGID1 with the corresponding residues of SmGID1s. Some replacements did not change the GA affinity or specificity, whereas other replacements diminished the binding activity (Supplementary Fig. 9 and Table 1). The replacement of Ile 133, which contacts the C2 in the A-ring (Figs 2b and 3a), with Leu (SmGID1a) or Val (SmGID1b) increased the binding affinity to GA$_{34}$ by several fold while diminishing the GA$_4$ binding affinity to less than one-quarter of that of the wild type (Table 1). The oxidation of C2 by the enzyme GA 2-oxidase has been considered an important reaction for the inactivation of active GAs[4], but no GA 2-oxidase-like sequences have been found in the *S. moellendorffii* genome (K. Hirano *et al.*, unpublished data). These observations indicate that the GA receptors in *S. moellendorffii* are less stringent in eliminating 2β-hydroxylated GAs and that this plant may not contain a deactivation system carried out by GA 2-oxidase. Furthermore, S127A showed much higher binding affinity for the non-3β-hydroxy GA (GA$_9$) than did the native OsGID1, indicating that Ser at this position is important for perceiving the 3β-hydroxyl group, the presence of which is required for growth-stimulatory activity in higher plants. SmGID1b contains Met at this position

Table 1 | The apparent K_d value of the wild-type and mutant GID1s for GA molecules

	Wild-type GID1	S127A	F27L	I133L	I133V	L330I
GA$_4$	5×10^{-7} (100)	2×10^{-6} (100)	5×10^{-7} (100)	2×10^{-6} (100)	3×10^{-6} (100)	5×10^{-7} (100)
GA$_9$	$>1 \times 10^{-4}$ (<0.5)	1×10^{-5} (20)	$>1 \times 10^{-4}$ (<0.5)	$>1 \times 10^{-4}$ (<2)	$>1 \times 10^{-4}$ (<3)	$>1 \times 10^{-4}$ (<0.5)
GA$_{34}$	$>1 \times 10^{-4}$ (<0.5)	ND (−)	$>1 \times 10^{-4}$ (<0.5)	3×10^{-5} (7)	5×10^{-5} (6)	$>1 \times 10^{-4}$ (<0.5)

All values are given as mol l^{-1}. The relative affinity to each mutated GID1, based on the value of GA$_4$, is shown in parentheses. ND, not determined.

and shows no difference in affinity for GA_4 and GA_9 (ref. 7). Although we could not directly measure the binding affinity of S127M (Supplementary Fig. 5) due to its low affinity for dihydro-GA_4, it is possible that the Met assigned at this position increases accessibility for non-3β-hydroxylated GAs. Together, these observations indicate that GID1 originated from a member of the HSLs, with its catalytic residue (Ser) having undergone adaptation for GA binding with both the carboxyl group at C6 of GA by means of a hydrogen bond and the helix dipole of α3, and some of the N-terminal lid residues having become specialized for holding the GA molecule by means of hydrophobic interactions. This was accomplished by replacement of some important amino acid residues that were gradually refined for high affinity and specificity to the currently active higher plant GAs such as GA_4 and GA_1.

METHODS SUMMARY

The recombinant proteins were purified with a His-tag affinity column (Bio-Rad) and were further purified by Superdex-200 gel filtration chromatography (GE Healthcare). The recombinant proteins were concentrated to 20 mg ml^{-1} in buffer containing 10 mM sodium phosphate (pH 7.5), 200 mM NaNO$_3$, 10 mM β-D-octylglucopyranoside, 1 mM DTT and 2 mM GA_4 for crystallization.

The GID1–GA_4 crystals were obtained in space group $P2_1$ ($a = 82.8$ Å, $b = 133.9$ Å, $c = 118.9$ Å, $\beta = 105.0°$) and contained six molecules designated chain A to F in the asymmetric unit. The structure was phased to 2.9 Å by single-wavelength anomalous dispersion from a mercury derivative, and traced using 1.9 Å native data. The final refined model, with an R factor of 20.1% and R_{free} of 24.4%, contains residues 14–353 of GID1 with some missing internal fragments. The GID1–GA_3 structure was solved by molecular replacement using the GID1–GA_4 structure as an initial model. The structural differences within the six molecules in the asymmetric unit are small. Thus, the chain A is used for the structural description.

The GA-binding assays and the yeast two-hybrid assays were performed according to the methods described in a previous report[2].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 September; accepted 14 October 2008.

1. Olszewski, N., Sun, T. P. & Gubler, F. Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell* **14** (Suppl.), S61–S80 (2002).
2. Ueguchi-Tanaka, M. *et al.* GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* **437**, 693–698 (2005).
3. Ueguchi-Tanaka, M. *et al.* Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. *Plant Cell* **19**, 2140–2155 (2007).
4. Davies, P. J. (ed.) *Plant Hormones: Biosynthesis, Signal Transduction, Action!* 3rd edn (Kluwer, 2004).
5. Ueguchi-Tanaka, M., Nakajima, M., Ashikari, M. & Matsuoka, M. Gibberellin receptor and its role in gibberellin signaling in plants. *Annu. Rev. Plant Biol.* **58**, 183–198 (2007).
6. Itoh, H., Ueguchi-Tanaka, M. & Matsuoka, M. Molecular biology of gibberellins signaling in higher plants. *Int. Rev. Cell Mol. Biol.* **268**, 191–221 (2008).
7. Hirano, K. *et al.* The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte *Selaginella moellendorffii* but not in the Bryophyte *Physcomitrella patens. Plant Cell* **19**, 3058–3079 (2007).
8. Yasumura, Y., Crumpton-Taylor, M., Fuentes, S. & Harberd, N. P. Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land–plant evolution. *Curr. Biol.* **17**, 1225–1230 (2007).
9. Yeaman, S. J. Hormone-sensitive lipase—new roles for an old enzyme. *Biochem. J.* **379**, 11–22 (2004).
10. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
11. Ileperuma, N. R. *et al.* High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon. *Biochemistry* **46**, 1851–1859 (2007).
12. Ollis, D. L. *et al.* The α/β hydrolase fold. *Protein Eng.* **5**, 197–211 (1992).
13. Holm, C., Davis, R. C., Osterlund, T., Schotz, M. C. & Fredrikson, G. Identification of the active site serine of hormone-sensitive lipase by site-directed mutagenesis. *FEBS Lett.* **344**, 234–238 (1994).

METHODS

Protein expression and purification. The full-length Os*GID1* complementary DNA sequences were inserted into pET11b vector (Novagen) with the His-tag coding sequence and overexpressed in *Esherichia coli* strain Rosetta (DE3) pLysS (Novagen). The recombinant proteins were purified as described in the text. For crystallization of the GID1–GA$_3$ complex, proteins were concentrated in buffer containing 50 mM sodium phosphate (pH 7.5), 200 mM NaNO$_3$, 10 mM β-D-octylglucopyranoside, 1 mM DTT and 10 mM GA$_3$.

Crystallization, data collection and structure determination. GID1–GA (GA$_4$ or GA$_3$) crystals were obtained at 20 °C by the sitting-drop vapour diffusion method with 1 μl protein solution mixed with an equal volume of reservoir solution containing 0.1 M HEPES (pH 7.5), 12–17% PEG 4000, 6–9% 2-methyl-2,4-pentanediol (MPD) and 0.1–0.2 M NaNO$_3$. Mercury (Hg) derivative crystals of GID1–GA$_4$ were prepared by soaking the native crystal in the reservoir solution supplemented with 0.025 mM thimerosal. For cryoprotection, the crystals were transferred into mother liquor containing 100 mM HEPES (pH 7.5), 20% PEG 4000, 10% MPD, 400 mM NaNO$_3$, 10 mM β-D-octylglucopyranoside and 2 mM gibberellin, and were flash-frozen in liquid nitrogen.

X-ray diffraction data were collected (90 K) at SPring-8 BL41XU (Harima, Japan). The crystals belonged to space group $P2_1$ with unit cell parameters $a = 82.8$ Å, $b = 133.9$ Å, $c = 118.9$ Å, $\beta = 105.0°$, and contained six molecules in the asymmetric unit. Data were integrated and scaled with HKL2000 (ref. 14; Supplementary Table 1).

The GID1–GA$_4$ structure was determined by the Hg single-wavelength anomalous diffraction (SAD) method. The 36 Hg sites were found and initial phases calculated to a resolution of 2.9 Å using PHENIX[15]; the model was built automatically using the native data to 2.0 Å resolution by ARP/wARP software[16].

Substantial model rebuilding was carried out manually with TURBO-FRODO[17] and COOT[18]. The overall structures were refined by Refmac5 (ref. 19; Supplementary Table 1). Molecular graphics (Figs 1 and 2 and Supplementary Figs 2, 3, 4 and 6) were illustrated with PyMOL (http://www.pymol.org).

GA binding and yeast two-hybrid assays. Mutations in OsGID1 proteins were introduced using PCR methods by designing mutated residues in primers. The mutated genes were constructed as described previously[3]. GA-binding assays were performed according to the methods described in a previous report[2] using [1,2,16,17-³H$_4$]-16,17-dihydro-GA4 (4.55 TBq mmol^{-1}), which was custom-ordered from DuPont/NEN. The median inhibitory concentration of each examined GA was calculated as for its apparent dissociation constant value, by a competitive binding assay with three or four concentrations of the GA.

The yeast two-hybrid assay was performed according to the methods described in a previous report[2].

14. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).

15. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

16. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).

17. Roussel, A. & Cambillau, C. *Silicon Graphics Geometry Partners Directory*, Vol. 81 77–78 (Silicon Graphics, Mountain View, 1991).

18. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

19. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

LETTERS

53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility

Nadya Dimitrova[1], Yi-Chun M. Chen[2], David L. Spector[2] & Titia de Lange[1]

Double-strand breaks activate the ataxia telangiectasia mutated (ATM) kinase, which promotes the accumulation of DNA damage factors in the chromatin surrounding the break. The functional significance of the resulting DNA damage foci is poorly understood. Here we show that 53BP1 (also known as TRP53BP1), a component of DNA damage foci, changes the dynamic behaviour of chromatin to promote DNA repair. We used conditional deletion of the shelterin component TRF2 (also known as TERF2) from mouse cells (TRF2^{fl/−}) to deprotect telomeres, which, like double-strand breaks, activate the ATM kinase, accumulate 53BP1 and are processed by non-homologous end joining (NHEJ)[1,2]. Deletion of TRF2 from 53BP1-deficient cells established that NHEJ of dysfunctional telomeres is strongly dependent on the binding of 53BP1 to damaged chromosome ends. To address the mechanism by which 53BP1 promotes NHEJ, we used time-lapse microscopy to measure telomere dynamics before and after their deprotection. Imaging showed that deprotected telomeres are more mobile and sample larger territories within the nucleus. This change in chromatin dynamics was dependent on 53BP1 and ATM but did not require a functional NHEJ pathway. We propose that the binding of 53BP1 near DNA breaks changes the dynamic behaviour of the local chromatin, thereby facilitating NHEJ repair reactions that involve distant sites, including joining of dysfunctional telomeres and AID (also known as AICDA)-induced breaks in immunoglobulin class-switch recombination.

Previous work has shown that mouse telomeres lacking TRF2 are processed by a KU70- and DNA-ligase-IV-dependent NHEJ reaction[1,2] that requires ATM kinase signalling and is stimulated by the ATM targets H2AX (also known as H2AFX) and MDC1 (refs 3 and 4). Here we focus on 53BP1, a third ATM target, which accumulates at double-strand breaks (DSBs) and deprotected telomeres[5–8]. The interaction of 53BP1 with chromatin involves the binding of its tudor domains to H4K20me2 and an MDC1-dependent interaction with γ-H2AX[9–14]. Although 53BP1 is not strictly required for DNA damage signalling, homology-directed repair or NHEJ in the context of V(D)J recombination, NHEJ of DSBs in class-switch recombination (CSR) is severely affected by 53BP1 deficiency[15,16]. In the absence of 53BP1, DSBs in different switch regions fail to join successfully, resulting in a predominance of intra-switch recombination events[17]. It has been proposed that 53BP1 might either facilitate synapsis of DNA ends[17,18] or 'shepherd' NHEJ factors to the break[19].

We generated SV40 LT immortalized TRF2^{fl/−}53BP1^{−/−} and TRF2^{fl/−}53BP1^{+/−} mouse embryonic fibroblasts (MEFs)[1,20] (Supplementary Fig. 1a, b) and assayed the frequency of telomere fusions in metaphase spreads collected 120 h after Cre-mediated deletion of TRF2 (Fig. 1a). Whereas 53BP1-proficient cells showed the expected level of telomere fusions (33% of telomeres fused after four population doublings), the rate of NHEJ in 53BP1^{−/−} cells was at least 50-fold lower, nearly as low as that in DNA-ligase-IV-defi-

cient cells[1,2]. The effect of 53BP1 on telomere fusions was also obvious when assayed by an in-gel hybridization assay (Fig. 1b). Cells deficient for 53BP1 failed to accumulate high-molecular-weight fusion products after deletion of TRF2 and showed no loss of the telomeric 3′ overhang, a second index for telomere fusion (Fig. 1b, c). In fact, the overhang signal increased, similar to what is observed when TRF2 is deleted from KU70- or DNA-ligase-IV-deficient cells[1,2]. In contrast to NHEJ, DNA damage signalling was not affected by the absence of 53BP1 (Fig. 1d, e). Deficiency of 53BP1 affected neither the phosphorylation of ATM and its target CHK2 (also known as CHEK2) nor the presence of γ-H2AX, MDC1 and NBS1 (also known as NBN) at dysfunctional telomeres (Fig. 1d, e and Supplementary Fig. 2a, b). The NHEJ defect was also not due to a change in cell cycle progression. After deletion of TRF2, the cells underwent the same number of cell divisions regardless of their 53BP1 status (Fig. 1f), and their S-phase index was not affected by 53BP1 deficiency (Supplementary Fig. 1c).

We next tested whether the tudor-domain-mediated recognition of H4K20me2 contributed to the promotion of NHEJ. TRF2^{fl/−}53BP1^{−/−} cells were complemented with wild-type human 53BP1 or with a tudor domain mutant (53BP1(D1521A)) with impaired binding to H4K20me2 (ref. 14; Fig. 2a). Although the two forms of 53BP1 were expressed at levels comparable to endogenous 53BP1 (Fig. 2b and data not shown), the recruitment of 53BP1(D1521A) to deprotected telomeres was diminished (Supplementary Fig. 3). In contrast to wild-type 53BP1, which reconstituted the NHEJ of dysfunctional telomeres, cells expressing the tudor mutant of 53BP1 showed a considerable delay in the fusion of telomeres lacking TRF2 (Fig. 2c, d).

The finding that the tudor domain mutation in 53BP1 did not abrogate NHEJ suggested that a second, H4K20me2-independent, interaction of 53BP1 with chromatin contributes to NHEJ. Because MDC1 is required for the stable association of 53BP1 with DSBs[9–14], we generated TRF2^{fl/fl}Mdc1^{−/−} MEFs[18] and determined their ability to execute the telomere fusion reaction. As expected, the binding of 53BP1 near deprotected telomeres was impaired by MDC1 deficiency, despite normal levels of H2AX phosphorylation (Supplementary Fig. 4a, b). MDC1 deficiency resulted in a delay in the NHEJ of telomeres but did not abrogate this process (Supplementary Fig. 5a, b), confirming our previous findings with MDC1 short hairpin RNAs (shRNAs)[3]. Furthermore, inhibition of MDC1 with shRNAs affected the residual NHEJ of telomeres in the context of the 53BP1 tudor domain mutant (Supplementary Fig. 5c–e).

These data establish an important role for chromatin-bound 53BP1 in the NHEJ of deprotected telomeres. Previous data showed that 53BP1 contributes to NHEJ in CSR[15,16], but not in other settings such as V(D)J recombination. We argued that the crucial difference between the 53BP1-dependent and-independent NHEJ reactions

[1]The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. [2]Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA.
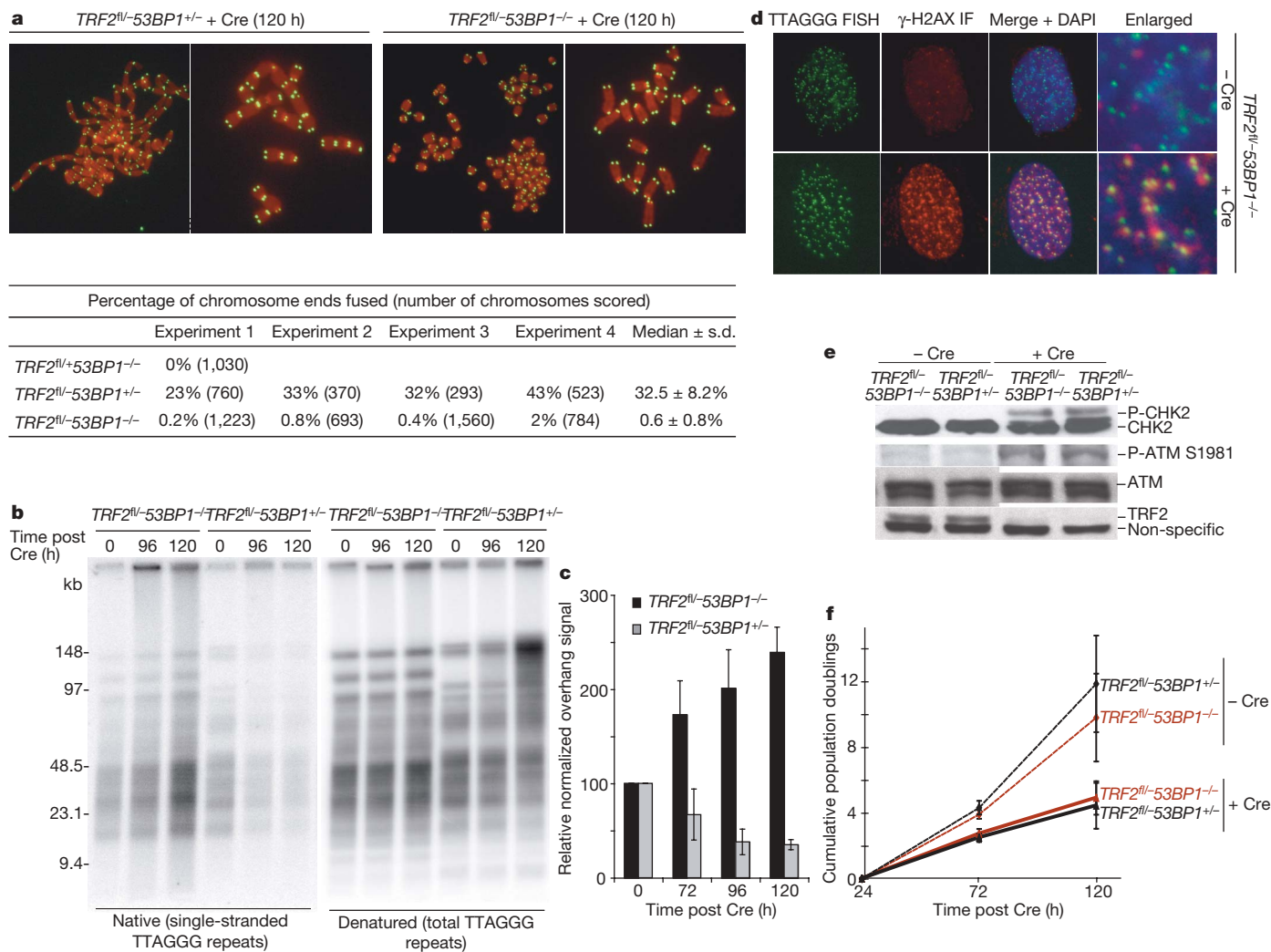
Figure 1 | Requirement for 53BP1 in NHEJ of dysfunctional telomeres, but not in DNA damage signalling. a, Metaphase chromosomes after deletion of TRF2 from the indicated cells (top). Telomeric fluorescence *in situ* hybridization (FISH), green; 4,6-diamidino-2-phenylindole (DAPI), red. Summary of the effect of 53BP1 on telomere fusions 120 h after TRF2 deletion (bottom). **b,** In-gel assay for the 3′ overhang (left) and total telomeric DNA (right) after deletion of TRF2 from 53BP1-proficient and -deficient cells. **c,** Quantification of overhang signals in **b** (mean ± s.d.; *n* = 3). **d,** Immunofluorescence (IF) for the presence of γ-H2AX at telomeres after deletion of TRF2 from 53BP1-proficent and -deficient cells. **e,** Immunoblots for phosphorylated (P-) ATM and CHK2 after TRF2 deletion. **f,** Proliferation of MEFs of the indicated genotype and treatment (mean ± s.d.; *n* = 3).

might be the distance between the DNA ends involved. Whereas the two ends generated by RAG1/2 or chromosome-internal DNA damage are close together, the DNA ends generated by AID in CSR are often far apart, as are dysfunctional telomeres in G1 when they are processed by NHEJ (ref. 21).

We used time-lapse microscopy to address whether 53BP1 altered the synapsis and/or dynamics of deprotected telomeres. To image telomeres, we used enhanced-green-fluorescent-protein-tagged version of the shelterin component TRF1 (eGFP–TRF1), which remains associated with telomeres when TRF2 is removed (Fig. 3a–c and Supplementary Fig. 6). Overexpression of this and other forms of TRF1 does not affect the protective function of telomeres[22] (data not shown). To mark sites of DNA damage signalling, we fused mCherry to amino acids 1220–1711 of 53BP1 (mCherry–BP1-2, see Figs 2a and 3a, b), creating a fusion protein that lacks most of the functional domains of 53BP1. On deletion of TRF2, mCherry–BP1-2 re-localized to telomeric sites containing eGFP–TRF1 (Fig. 3c). The mCherry–BP1-2 allele did not affect the frequency of telomere fusions in 53BP1-proficient cells nor did it restore the NHEJ defect in 53BP1-deficient cells (Supplementary Fig. 7a), establishing that it is a neutral marker for DNA damage in this context. Furthermore,

imaging experiments repeated in cells lacking mCherry–BP1-2 gave the same outcome (Supplementary Fig. 7b, c and see below).

To analyse the effect of 53BP1 on the movement of dysfunctional telomeres, time-lapse microscopy was performed at an early stage after deletion of TRF2 (72–84 h post-Cre) when most TRF1-marked sites still represent free chromosome ends[3]. Initial analysis of eGFP and mCherry signals indicated that dysfunctional telomeres became more mobile but only when 53BP1 was present (Supplementary Videos 1 and 2). We observed occasional apparent fusion events in 53BP1-proficient cells (Supplementary Video 3), but no potential fusions were observed in 53BP1-deficient cells, even for telomeres that were closely apposed (Supplementary Video 4).

To obtain a quantitative measure of telomere mobility, individual eGFP–TRF1/mCherry–BP1-2-labelled dysfunctional telomeres were traced in deconvolved and projected images (Fig. 3a, d and Supplementary Fig. 8a, b). In parallel, eGFP–TRF1-marked functional telomeres were analysed in cells not treated with Cre. Functional telomeres in 53BP1-proficient and -deficient cells travelled in a constrained random walk over a total path of approximately 3.7–3.9 ± 0.7 μm at a median speed of 180–190 nm min⁻¹, which is comparable to the movement of human telomeres (Fig. 3e, f and
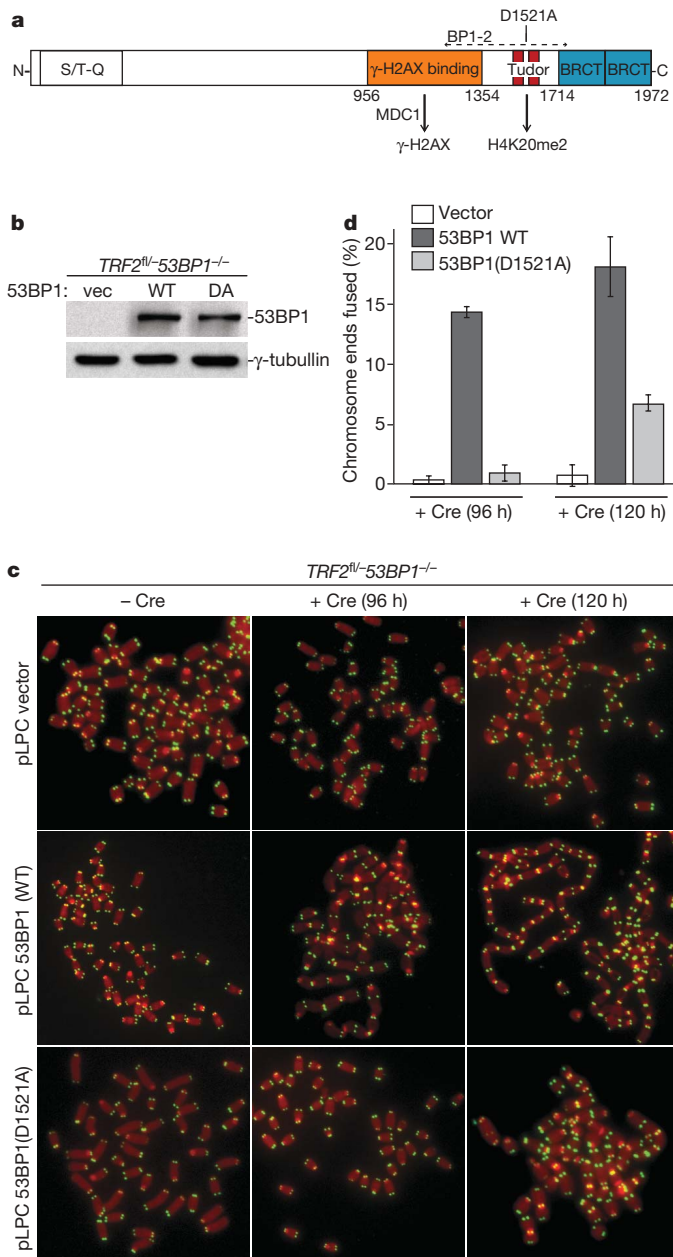
Figure 2 | Optimal NHEJ of dysfunctional telomeres requires interaction of 53BP1 with H4K20me2. a, Schematic of the domain structure of 53BP1 highlighting the N-terminal cluster of S/T-Q phosphorylation sites, the tudor domain and BRCT (BRCA1 C-terminal) repeats. **b**, Immunoblot for 53BP1 expression in $TRF2^{fl/-}53BP1^{-/-}$ cells complemented with empty vector (vec), wild type (WT) or the tudor mutant 53BP1(D1521A) (DA). **c**, Telomere fusions in metaphase spreads of $TRF2^{fl/-}53BP1^{-/-}$ MEFs complemented as indicated and treated with Cre as specified. **d**, Quantification of telomere fusions in **c** (mean ± s.d.; $n = 3$).

Supplementary Figs 7b and 9a)[23]. In contrast, dysfunctional telomeres were significantly more mobile, travelling at a speed of 270–360 nm min^{-1} over a cumulative distance ranging from 5.4 μm to 7.2 μm (Fig. 3e, f and Supplementary Figs 7b and 9a). The mobility increase associated with telomere dysfunction was significantly attenuated in cells lacking 53BP1 (Fig. 3e), resulting in a median cumulative distance travelled of 4.4 ± 0.2 μm. The presence of the mCherry–BP1-2 DNA damage marker did not affect the outcome; the same results were obtained after deletion of TRF2 from $53BP1^{+/-}$ or $53BP1^{-/-}$ cells that only contained the eGFP–TRF1 marker (Supplementary Fig. 7b, c). Importantly, in both settings the dysfunctional telomeres in 53BP1-proficient cells sampled larger

territories than functional telomeres (maximal displacement 1.2 ± 0.3 μm versus 0.51 ± 0.29 μm; Fig. 3g and Supplementary Figs 7c and 9b). Assuming that telomeres sample spheres within the nucleus, a twofold change in the radius of the two-dimensional territory will increase the volume sampled by a dysfunctional telomere by a factor of eight. Furthermore, a substantial fraction (>10%) of the dysfunctional telomeres roamed well beyond 2 μm, whereas none of the 115 functional telomeres analysed moved beyond that distance (Fig. 3g and Supplementary Fig. 7c). When 53BP1 was not present, the median maximal displacement of the dysfunctional telomeres was only 0.8 ± 0.2 μm, indicating movement within a significantly smaller territory, and only one out of 115 dysfunctional telomeres sampled an area beyond 2 μm (Fig. 3g and Supplementary Fig. 7c). These data establish that telomeres become more mobile and sample larger territories when they are deprived of their normal protection and that this change in their dynamic behaviour is promoted by 53BP1. Assuming that the rate of NHEJ correlates with the probability of an encounter between two telomeres, the ability of 53BP1 to expand the three-dimensional space visited by dysfunctional telomeres is a probable explanation for its effect on telomere fusion. We confirmed that the absence of the NHEJ reaction itself could not account for the slower movement because telomeres in DNA-ligase-IV-deficient cells still showed a considerable increase in their mobility on TRF2 deletion (Fig. 3h and Supplementary Fig. 9c, d).

The ability of 53BP1 to promote the mobility of dysfunctional telomeres could explain why ATM kinase, its upstream regulator, is required for telomere fusions[4]. Indeed, similar to 53BP1-deficient cells, dysfunctional telomeres failed to gain their maximal mobility in ATM-deficient cells (Fig. 3i and Supplementary Fig. 9c, d).

Next we tested whether 53BP1 caused global chromatin mobility in response to DNA damage. We tracked functional telomeres in wild-type or $Lig4^{-/-}$ cells immediately after treatment with γ-irradiation or after allowing the cells to recover. In all cases, we found that the mobility of the telomeres was unaffected by the induction of damage elsewhere in the genome (Supplementary Fig. 10), arguing that increased mobility is a local event, taking place at the site of damage where 53BP1 accumulates.

Because 53BP1 is the first DNA damage factor to be implicated in the movement of sites of DNA damage, it will be important to determine the mechanism by which it affects chromatin mobility. Latrunculin A did not alter the gain in mobility of deprotected telomeres (Supplementary Fig. 11a), arguing against an actin-driven process. Similarly, the histone deacetylase (HDAC) inhibitor trichostatin A had no effect on the mobility of the deprotected telomeres, making it unlikely that 53BP1 acts through its interaction with HDAC4 (ref. 24; Supplementary Fig. 11b). We imagine that 53BP1 dislodges a factor that restricts chromatin mobility or that 53BP1 endows the chromatin with dynamic features, for instance by altering higher order interactions that affect the length of thermal persistence.

These data reveal that the DNA damage factor 53BP1 is crucial for the NHEJ reaction at dysfunctional telomeres and suggest a mechanism by which 53BP1 promotes telomere fusions. Telomeres, like other chromosomal sites, have a limited range of motion, showing constrained diffusion within a territory with a radius of ≤0.5 μm (this work and refs 23 and 25). When telomeres become damaged, their mobility increases significantly, which is expected to increase the chance that two dysfunctional telomeres become closely apposed, thereby allowing the NHEJ machinery to engage the two DNA ends. Given that 53BP1 is also important for CSR, we suggest that the change in chromatin dynamics also promotes the joining of AID-induced breaks. In an accompanying paper[26], this paradigm is further extended to distant DSBs in V(D)J recombination. We therefore suggest that 53BP1 is a facilitator of NHEJ at all distant DNA ends, acting by increasing the mobility of the local chromatin. Because joining of distant DNA ends promotes chromosome rearrangements, we consider it unlikely that the main role of 53BP1 is to promote
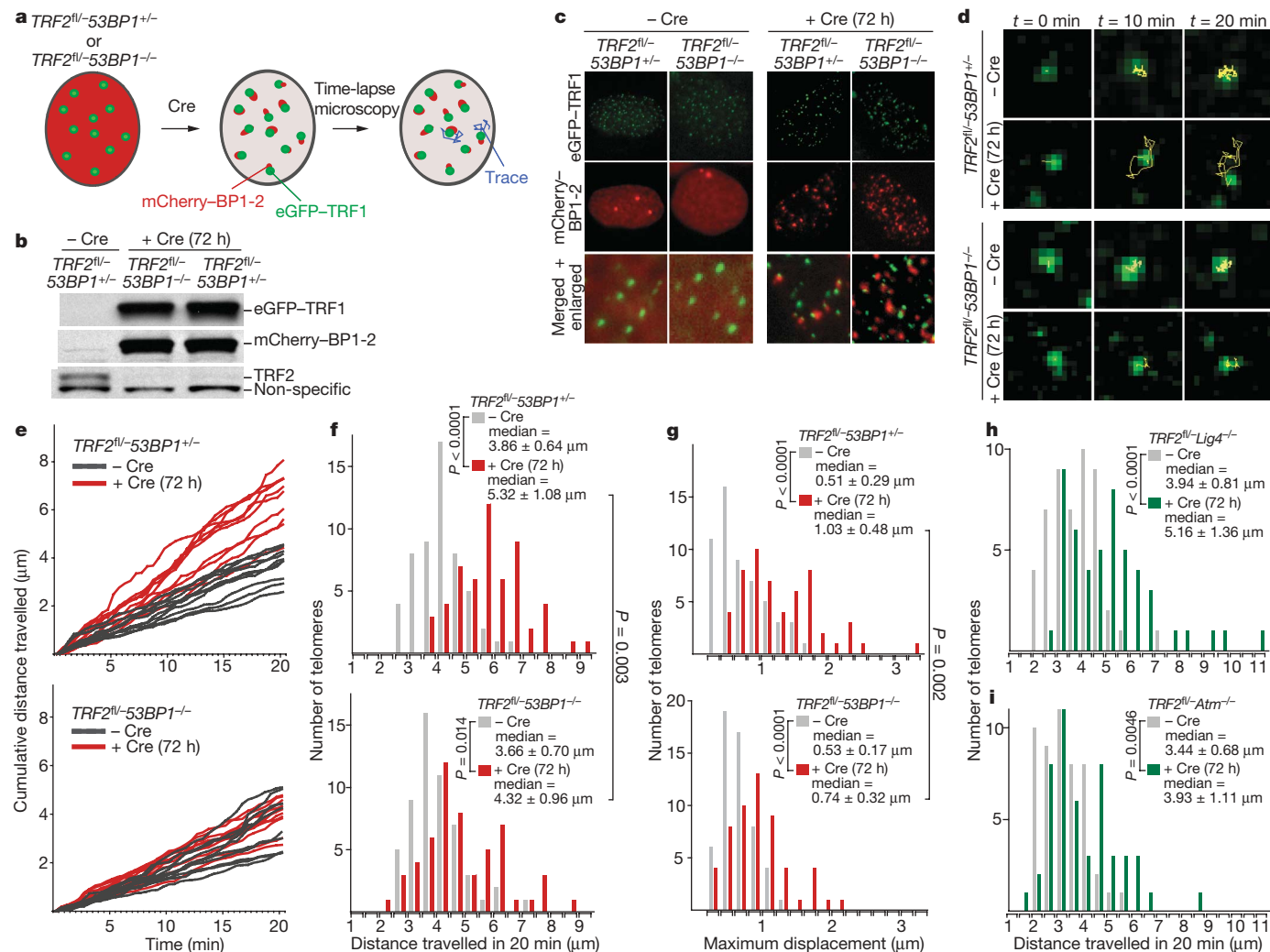
Figure 3 | 53BP1 promotes mobility of dysfunctional telomeres.
a, Schematic of the live-cell imaging experiments. **b**, Immunoblotting for eGFP–TRF1, mCherry–BP1-2 and TRF2. **c**, Co-localization of eGFP–TRF1 and mCherry–BP1-2 fluorescence signals. **d**, Representative traces of telomeres in indicated MEFs expressing eGFP–TRF1 and mCherry–BP1-2. Snapshots taken at indicated time points. **e**, Ten representative telomeres from indicated MEFs, treated as specified, were tracked as in **d**; the cumulative distance at each time point is plotted against time. **f**, Frequency distribution of the cumulative distance travelled in one representative experiment in MEFs of the indicated genotype and treatment (median ± s.d.; $n = 55$; P was calculated using a two-tailed Mann–Whitney test). **g**, Frequency distribution of the maximum displacement from the starting point registered by individual telomeres during the imaging session described in **f** (median ± s.d. and P values were calculated as in **f**). **h, i**, Analysis as in **f** in MEFs of the indicated genotypes and treatments, expressing the eGFP–TRF1 marker only ($n = 50$).

NHEJ at DNA ends that are far apart. It seems more likely that the increased chromatin mobility we observe reflects an attribute of 53BP1-containing chromatin that is important for DNA repair at all DSBs, including those that are closely apposed and breaks that are processed by homology-directed repair. We imagine that at telomeres this chromatin change leads to increased mobility, whereas at chromosome-internal DSBs the same change in the chromatin may facilitate their repair in other ways.

METHODS SUMMARY

MEFs from embryonic day (E)13.5 embryos obtained from crosses between $TRF2^{fl/-}$ and $53BP1^{+/-}$ mice[1,20] were isolated and immortalized at passage 2 with pBabeSV40LT (a gift from G. Hannon). Cre was introduced by retroviral infection with pMMP Hit&Run Cre[27]. For time-lapse microscopy, $TRF2^{fl/-}53BP1^{+/-}$ and $TRF2^{fl/-}53BP1^{-/-}$ cells expressing eGFP−TRF1 and mCherry−BP1-2, or $TRF2^{fl/-}Lig4^{-/-}p53^{-/-}$ (ref. 1) and $TRF2^{fl/-}Atm^{-/-}$ (ref. 4) cells expressing eGFP−TRF1 only, were imaged untreated or 72 h after Cre-mediated deletion of TRF2. Time-lapse images were acquired using a DeltaVision RT microscope system (Applied Precision) in three dimensions in both eGFP and mCherry channels using SoftWoRx software every 30 s for 20 min. Images were deconvolved, projected in two dimensions, and tracking analysis was performed with ImageJ. Data were collected from three independent experiments ($n \geq 50$) and Mann–Whitney statistical analysis was performed using Prism software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Celli, G. & de Lange, T. DNA processing not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biol.* **7**, 712–718 (2005).
2. Celli, G. B., Lazzerini Denchi, E. & de Lange, T. Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nature Cell Biol.* **8**, 855–890 (2006).
3. Dimitrova, N. & de Lange, T. MDC1 accelerates non-homologous end-joining of dysfunctional telomeres. *Genes Dev.* **20**, 3238–3243 (2006).
4. Lazzerini Denchi, E. & de Lange, T. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* **448**, 1068–1071 (2007).
5. DiTullio, R. A. Jr *et al.* 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. *Nature Cell Biol.* **4**, 998–1002 (2002).
6. Wang, B., Matsuoka, S., Carpenter, P. B. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science* **298**, 1435–1438 (2002).
7. Fernandez-Capetillo, O., Liebe, B., Scherthan, H. & Nussenzweig, A. H2AX regulates meiotic telomere clustering. *J. Cell Biol.* **163**, 15–20 (2003).

8. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).

9. Goldberg, M. *et al.* MDC1 is required for the intra-S-phase DNA damage checkpoint. *Nature* **421**, 952–956 (2003).

10. Lou, Z., Minter-Dykhouse, K., Wu, X. & Chen, J. MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways. *Nature* **421**, 957–961 (2003).

11. Stewart, G. S., Wang, B., Bignell, C. R., Taylor, A. M. & Elledge, S. J. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* **421**, 961–966 (2003).

12. Ward, I. M., Minn, K., Jorda, K. G. & Chen, J. Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. *J. Biol. Chem.* **278**, 19579–19582 (2003).

13. Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *J. Cell Biol.* **170**, 201–211 (2005).

14. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).

15. Manis, J. P. *et al.* 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nature Immunol.* **5**, 481–487 (2004).

16. Ward, I. M. *et al.* 53BP1 is required for class switch recombination. *J. Cell Biol.* **165**, 459–464 (2004).

17. Reina-San-Martin, B., Chen, J., Nussenzweig, A. & Nussenzweig, M. C. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1$^{-/-}$ B cells. *Eur. J. Immunol.* **37**, 235–239 (2007).

18. Lou, Z. *et al.* MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* **21**, 187–200 (2006).

19. Xie, A. *et al.* Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. *Mol. Cell* **28**, 1045–1057 (2007).

20. Ward, I. M., Minn, K., van Deursen, J. & Chen, J. p53 binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. *Mol. Cell. Biol.* **23**, 2556–2563 (2003).

21. Konishi, A. & de Lange, T. Cell cycle control of telomere protection and NHEJ revealed by a ts mutation in the DNA-binding domain of TRF2. *Genes Dev.* **22**, 1221–1230 (2008).

22. van Steensel, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).

23. Molenaar, C. *et al.* Visualizing telomere dynamics in living mammalian cells using PNA probes. *EMBO J.* **22**, 6631–6641 (2003).

24. Kao, G. D. *et al.* Histone deacetylase 4 interacts with 53BP1 to mediate the DNA damage response. *J. Cell Biol.* **160**, 1017–1027 (2003).

25. Abney, J. R., Cutler, B., Fillbach, M. L., Axelrod, D. & Scalettar, B. A. Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *J. Cell Biol.* **137**, 1459–1468 (1997).

26. Difilippantonio, S. *et al.* 53BP1 Facilitates long-range DNA end-joining during V(D)J recombination. *Nature* doi:10.1038/nature07476 (this issue).

27. Silver, D. P. & Livingston, D. M. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol. Cell* **8**, 233–243 (2001).

METHODS

Generation of MEFs and deletion of TRF2. MEFs from E13.5 embryos obtained from crosses between *TRF2*[fl/−] and *53BP1*[+/−] mice[1,20] and between *TRF2*[fl/fl] and *Mdc1*[+/−] mice[18] were isolated and immortalized at passage 2 with pBabeSV40LT. To delete TRF2, Cre was introduced by retroviral infection using the pMMP Hit&Run Cre retroviral construct[27]. In brief, Hit&Run Cre was expressed in ecotrophic Phoenix cells. Virus-containing supernatant was collected at 36, 48 and 60 h post-transfection and MEFs were infected consecutively three times every 12 h. The medium was changed 12 h after the last infection, and cells were analysed at the indicated time points after the second infection.

Constructs. Full-length human 53BP1 was cloned by PCR into a N-Myc-pLPC-puro retroviral expression vector. The D1521A mutation was introduced by a PCR-based mutagenesis strategy using the following mutagenesis primers: 5′-AAATTGCTCTTTGATGCTGGGTACGAATGTGAT-3′ and 5′-ATCACATTCGTACCCAGCATCAAAGAGCAATTT-3′. The mutation was confirmed by sequencing. Wild-type and D1521A rescue alleles were introduced into *TRF2*[fl/−] *53BP1*[−/−] cells by five consecutive retroviral infections, delivered at 12-h intervals with virus-containing supernatants from Phoenix cells. Puromycin selection was applied. Infection with the empty vector was used as a negative control.

Fluorescently amino-terminally tagged mCherry−BP1-2 and eGFP−TRF1 were cloned by PCR into pLPC-puro and pWzl-hygro, respectively. In-frame fusions were confirmed by sequencing. mCherry−BP1-2 and eGFP−TRF1 constructs were consecutively introduced into *TRF2*[fl/−] *53BP1*[+/−] and *TRF2*[fl/−] *53BP1*[−/−] cells by retroviral infections, followed by puromycin (4 days) and hygromycin (7 days) selection, respectively.

MDC1 shRNAs. MDC1 was stably reduced in *TRF2*[fl/−] *53BP1*[−/−] cells expressing the 53BP1(D1521A) mutant using shRNAs expressed from the pSUPERIOR retroviral vector (OligoEngine). Retrovirus was produced in ecotrophic Phoenix cells and used to infect cells four times at 12-h intervals. No selection was applied. The following target sequences[3] were cloned into pSUPERIOR and confirmed by DNA sequencing: mouse *Mdc1* sh4, 5′-ACAGCATGCAGTAATTGAA-3′; mouse *Mdc1* sh5, 5′-ACACAGCCGTTCTGTCTAA-3′. Infections with the empty vector were used as a negative control. Efficient knockdown of *Mdc1* was verified by immunoblotting.

Telomere fluorescence *in situ* hybridization. Cells were collected at the indicated time points after Cre infection and fixed as described previously[28]. Metaphase spreads were aged overnight and peptide nucleic acid (PNA) FISH[29] was performed. In brief, slides were washed in PBS once and fixed in 4% formaldehyde for 2 min at room temperature (24–25 °C). After extensive PBS washes, spreads were digested for 10 min at 37 °C with 1 mg ml^{-1} pepsin dissolved in 10 mM glycine, pH 2.2. Slides were then washed in PBS, fixed again in 4% formaldehyde for 2 min at room temperature, and washed in PBS before dehydration by consecutive 5-min incubations in 70%, 95% and 100% ethanol. After air-drying, Hybridizing Solution (70% formamide, 1 mg ml^{-1} blocking reagent (Roche), 10 mM Tris-HCl, pH 7.2) containing FIu-OO-(AATCCC)$_3$ PNA probe (Applied Biosystems) was added and spreads were denatured by heating for 3 min at 80 °C on a heat block. Spreads were then allowed to hybridize in the dark for 2 h at room temperature. Two 15-min washes were performed in a mixture containing 70% formamide, 10 mM Tris-HCl, pH 7.0, and 0.1% BSA, followed by three washes in a mixture containing 0.1 M Tris-HCl, pH 7.0, 0.15 M NaCl and 0.08% Tween-20, with DAPI added to the second wash to counter-stain the chromosomal DNA. Slides were mounted in antifade reagent (ProLong Gold, Invitrogen), and digital images were captured with a Zeiss Axioplan II microscope with a Hamamatsu C4742-95 camera using Improvision OpenLab software.

In-gel analysis of telomeric DNA. Pulse-field gel electrophoresis and in-gel detection of mouse telomeric DNA were performed as described previously[1]. Cells were resuspended in PBS and mixed 1:1 (v/v) with 2% agarose (SeaKem agarose) to obtain 5×10^5 cells per agarose plug. Plugs were digested overnight with 1 mg ml^{-1} Proteinase K (in buffer containing 100 mM EDTA, 0.2% sodium deoxycolate, 1% sodium lauryl sarcosine), washed extensively with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and incubated overnight at 37 °C with 60 U MboI. The following day, the plugs were washed once in TE and once in water, and were equilibrated in 0.5 × TBE. Plugs were loaded on a 1% agarose/0.5 × TBE gel and run for 24 h using CHEF-DRII PFGE apparatus (BioRad) in 0.5 × TBE running buffer. The settings were as follows: initial pulse, 5 min; final pulse, 5 min; 6 V cm^{-1}; 14 °C. In-gel hybridization of the native gel with ^{32}P-γATP end-labelled (CCCTAA)$_4$ oligonucleotides and subsequent denaturation and hybridization steps were performed as described[30]. Gels were exposed onto a PhosphoImager screen overnight, and the single-stranded G-overhang signal was quantified with ImageQuant software and normalized to the total telomeric DNA quantified after denaturation. The percentage overhang value given repre-

sents the percentage of overhang signal detectable at different time points compared to the overhang signal for the same cells not treated with Cre.

Immunoblotting. Cells were lysed in 2 × Laemmli buffer (100 mM Tris-HCl, pH 6.8, 200 mM DTT, 3% SDS, 20% glycerol, 0.05% bromophenol blue) at 10^4 cells per microlitre, denatured for 7 min at 100 °C, and sheared with an insulin needle before loading the equivalent of 2×10^5 cells per lane. After immunoblotting, membranes were blocked in PBS with 5% milk and 0.1% Tween, and incubated with the following primary antibodies in 5% milk and 0.1% Tween: 53BP1 (rabbit polyclonal; 100-305, Novus Biologicals; detects endogenous mouse 53BP1, full-length wild-type and mutant human 53BP1 rescue alleles, and mCherry−BP1-2 fusion protein); MDC1 (rabbit polyclonal; 300-757, Bethyl); TRF2 (rabbit polyclonal; 647 (ref. 31)); TRF1 (1449, rabbit polyclonal; A. Sfeir and T.d.L., unpublished); CHK2 (mouse monoclonal, BD Biosciences); ATM S1981-P (mouse monoclonal; 10H11.E12, Cell Signaling); or ATM (mouse monoclonal; clone MAT3, Sigma). γ-Tubulin (mouse monoclonal, clone GTU 488, Sigma) or a non-specific band in the TRF2 western blot were used as a loading control. Blots were developed with enhanced chemiluminescence (Amersham).

Immunofluorescence-FISH. Cells were grown on coverslips and fixed for 10 min in 2% paraformaldehyde at room temperature followed by PBS washes (IF for γ-H2AX, MDC1 and 53BP1) or fixed for 10 min in methanol:acetone (1:1) at −20 °C followed by rehydration in PBS for 5 min (IF for NBS1). Coverslips were blocked for 30 min in blocking solution (1 mg ml^{-1} BSA, 3% goat serum, 0.1% Triton X-100, 1 mM EDTA in PBS). Next, the cells were incubated with the following primary antibodies diluted in blocking solution for 1 h at room temperature: 53BP1 (rabbit polyclonal; 100-304A, Novus Biologicals); γ-H2AX-S139-P (mouse monoclonal; Upstate Biotechnology); MDC1 (mouse monoclonal, a gift from J. Chen); and NBS1 (rabbit polyclonal, a gift from J. Petrini). After PBS washes, coverslips were incubated with Rhodamine-Red-X-labelled secondary antibody raised against mouse or rabbit (RRX, Jackson ImmunoResearch) for 30 min and washed in PBS. At this point, coverslips were fixed with 2% paraformaldehyde for 10 min at room temperature, washed extensively in PBS, dehydrated consecutively in 70%, 95% and 100% ethanol for 5 min each, and allowed to dry completely. Hybridizing solution (70% formamide, 1 mg ml^{-1} blocking reagent (Roche), 10 mM Tris-HCl, pH 7.2, containing PNA probe FITC-OO-(AATCCC)$_3$ (Applied Biosystems)) was added to each coverslip and the cells were denatured by heating for 10 min at 80 °C on a heat block. After 2 h incubation at room temperature in the dark, cells were washed twice with washing solution (70% formamide, 10 mM Tris-HCl, pH 7.2) and twice in PBS. DNA was counterstained with DAPI and slides were mounted in antifade reagent (ProLong Gold, Invitrogen). Digital images were captured with a Zeiss Axioplan II microscope with a Hamamatsu C4742-95 camera using Improvision OpenLab software.

To detect expression and localization of fluorescently marked proteins, cells were fixed for 10 min in 2% paraformadehyde at room temperature. Digital images of fluorescent eGFP and mCherry signals were captured as described previously.

Live-cell imaging. *TRF2*[fl/−] *53BP1*[+/−] and *TRF2*[fl/−] *53BP1*[−/−] cells expressing eGFP−TRF1 (to visualize telomeres) and mCherry−BP1-2 (human 53BP1, amino acids 1220−1711, to mark dysfunctional telomeres), or *TRF2*[fl/−] *53BP1*[+/−], *TRF2*[fl/−] *53BP1*[−/−], *TRF2*[fl/−] *Lig4*[−/−] *p53*[−/−] (ref. 1) and *TRF2*[fl/−] *Atm*[−/−] (ref. 4) cells expressing eGFP−TRF1 only, untreated or treated with Cre, were seeded onto MatTek glass bottom plates and grown for 2 days before imaging. Imaging was performed 72−84 h after Cre-mediated deletion of TRF2. Right before imaging, cells were changed into Leibovitz's L-15 medium (Gibco) supplemented with 30% FBS and allowed to equilibrate for 30 min. During the imaging session, the temperature was maintained at 37 °C with an environmental chamber. Cells were monitored using a DeltaVision RT microscope system (Applied Precision) with a PlanApo ×60 1.40 n.a. objective lens (Olympus America, Inc.). 5-μm Z-stacks at 0.5-μm steps in both eGFP and mCherry channels were acquired using SoftWoRx software with 50 ms and 30 ms exposure time, respectively, every 30 s over 20 min ($t = 40$ frames) at 2×2 binning with 512×512 pixels in final size. Images were deconvolved and projected in two dimensions using SoftWoRx software.

The tracking analysis of eGFP–TRF1-marked telomeres was performed with ImageJ software for at least ten cells for each genotype. Cells were registered by StackReg plugin using both 'Translation' and 'Scaled rotation' options. Next, particles were tracked using Particle Detector and Tracker plugin[32] with the following parameters for particle detection and tracking: radius = 2 pixels; cut-off = 2 pixels; percentile = 1; link range = 1; displacement = 5 pixels. The *x* and *y* coordinates of each trajectory were output for further calculation.

For each cell, 5 telomeres were chosen for analysis based on two criteria: first, they were continuously tracked for at least 35 out of 40 frames and, second, they co-localized with the mCherry–BP1-2 dysfunctional telomere marker for the

entire imaging session (20 min). To correct for cell mobility, the average x and y values of the 5 telomeres were calculated in each frame and this was used as a reference point. All data output in pixels (standard ImageJ output) were converted to metres by the formula 1 pixel = 0.2156 µm based on the characteristics of the objective.

The following formulas were used to calculate the distance travelled between two time points, cumulative distance travelled and average speed of an individual telomere T ($x_T^{t=n}$, $y_T^{t=n}$) relative to the reference point R ($x_R^{t=n}$, $y_R^{t=n}$): displacement, D_n, between two time points $t = n-1$ and $t = n$ (measured in µm) is

$$D_n = \mathrm{sqrt}(((x_T^{t=n} - x_R^{t=n}) - (x_T^{t=n-1} - x_R^{t=n-1}))^2 + ((y_T^{t=n} - y_R^{t=n}) - (y_T^{t=n-1} - y_R^{t=n-1}))^2)$$

Cumulative distance travelled in 20 min ($t = 40$) D_{cum} (measured in µm) is

$$D_{cum} = \mathrm{sum}(D_1, D_2, \ldots, D_{40})$$

Average speed S (measured in $\mathrm{nm\,min}^{-1}$) is

$$S = D_{cum}/20$$

To calculate the displacement from the starting point ($t = 0$) for a given telomere T ($x_T^{t=n}$, $y_T^{t=n}$) at $t = n$, the following calculation was performed based on a reference point, R, defined as above: displacement from origin $D_{ori,t}$ is

$$D_{ori,t} = \mathrm{sqrt}(((x_T^{t=n} - x_R^{t=n}) - (x_T^{t=0} - x_R^{t=0}))^2 + ((y_T^{t=n} - y_R^{t=n}) - (y_T^{t=0} - y_R^{t=0}))^2)$$

Maximum displacement from starting point, $D_{ori\ MAX}$, for a given telomere recorded during an imaging session was used as a measure of the territory that the telomere has sampled during the imaging session and calculated as shown below:

$$D_{ori\ MAX} = \max(D_{ori,\ 1}, D_{ori,\ 2}, \ldots, D_{ori,\ 40})$$

Data were collected from three independent experiments (independent Cre infections): experiment 1 ($n = 55$; Fig. 3e–g), experiment 2 ($n = 55$) and experiment 3 ($n = 65$; Supplementary Fig. 9). It is important to note that experiment 1 and experiment 3 were done with cell lines A ($TRF2^{fl/-}\,53BP1^{+/-}$) and B ($TRF2^{fl/-}\,53BP1^{-/-}$), whereas experiment 2 was performed with cell lines C ($TRF2^{fl/-}\,53BP1^{+/-}$) and D ($TRF2^{fl/-}\,53BP1^{-/-}$). A/B and C/D were infected with eGFP–TRF1/mCherry–BP1-2 fluorescent markers independently, which might account for the minor variation in the quantitative data on telomere mobility after deprotection.

Statistical analysis was performed using Prism Software. Mann–Whitney test (also referred to as rank sum test), which compares two unpaired groups without assuming Gaussian distribution, was applied to calculate the statistical significance values.

Treatment with drugs. Cells were treated with the following drugs diluted in imaging medium: actin inhibitor Latrunculin A, $0.1\,\mu\mathrm{g\,ml}^{-1}$, for 1 h before imaging and HDAC inhibitor trichostatin A, $10\,\mathrm{ng\,ml}^{-1}$ and $50\,\mathrm{ng\,ml}^{-1}$, for 18 h before imaging. $TRF2^{fl/-}$ or $TRF2^{fl/-}\,Lig4^{-/-}\,p53^{-/-}$ cells plated on imaging plates were irradiated with 1 Gy γ-irradiation and imaged immediately (with 10 min delay to set up the imaging) or allowed to recover for 2 h.

28. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
29. Lansdorp, P. M. *et al.* Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* **5**, 685–691 (1996).
30. Karlseder, J., Smogorzewska, A. & de Lange, T. Senescence induced by altered telomere state, not telomere loss. *Science* **295**, 2446–2449 (2002).
31. Zhu, X. D. *et al.* Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nature Genet.* **25**, 347–352 (2000).
32. Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195 (2005).

LETTERS

53BP1 facilitates long-range DNA end-joining during V(D)J recombination

Simone Difilippantonio[1], Eric Gapud[2], Nancy Wong[1], Ching-Yu Huang[2], Grace Mahowald[2], Hua Tang Chen[1], Michael J. Kruhlak[1], Elsa Callen[1], Ferenc Livak[3], Michel C. Nussenzweig[4]*, Barry P. Sleckman[2]* & André Nussenzweig[1]*

Variable, diversity and joining (V(D)J) recombination and class-switch recombination use overlapping but distinct non-homologous end joining pathways to repair DNA double-strand-break intermediates. 53BP1 is a DNA-damage-response protein that is rapidly recruited to sites of chromosomal double-strand breaks, where it seems to function in a subset of ataxia telangiectasia mutated (ATM) kinase-, H2A histone family member X (H2AX, also known as H2AFX)- and mediator of DNA damage checkpoint 1 (MDC1)-dependent events[1,2]. A 53BP1-dependent end-joining pathway has been described that is dispensable for V(D)J recombination but essential for class-switch recombination[3,4]. Here we report a previously unrecognized defect in the joining phase of V(D)J recombination in 53BP1-deficient lymphocytes that is distinct from that found in classical non-homologous-end-joining-, H2ax-, Mdc1- and Atm-deficient mice. Absence of 53BP1 leads to impairment of distal V–DJ joining with extensive degradation of unrepaired coding ends and episomal signal joint reintegration at V(D)J junctions. This results in apoptosis, loss of T-cell receptor α locus integrity and lymphopenia. Further impairment of the apoptotic checkpoint causes propagation of lymphocytes that have antigen receptor breaks. These data suggest a more general role for 53BP1 in maintaining genomic stability during long-range joining of DNA breaks.

Recombination-activating gene 1 (RAG1) and RAG2 (RAG1/2) proteins perform the pairwise cleavage step in V(D)J recombination, whereas activation-induced cytidine deaminase (AID) triggers the formation of double-strand breaks (DSBs) in the switch regions during class-switch recombination (CSR)[5]. RAG1/2- and AID-induced lesions in antigen receptor loci initiate nuclear focus formation of the DNA damage response proteins γ-H2AX, NBS1 (also known as NBN) and 53BP1 (TRP53BP1) over a large chromosome domain[5–8]. On the basis of the analysis of CSR in H2ax-deficient[9], Atm-deficient[7] and 53BP1-deficient[3,4,10,11] lymphocytes, it was proposed that focus-forming factors might promote and/or maintain synapsis of distal switch regions[9]. In the case of 53BP1 deficiency, there is an almost complete loss of long-range CSR[3,4,12] and a concomitant increase in the frequency of short-range intra-switch recombination[12]. Paradoxically, although V(D)J recombination and CSR use similar DSB repair mechanisms, there are no known defects in V(D)J recombination in the absence of 53BP1 (refs 3, 4).

$53BP1^{-/-}$ mice show a 50–80% reduction in the number of B and T lineage cells in bone marrow and thymus, respectively (Supplementary Fig. 1a, b and ref. 13). This is similar to the lymphopenia found in $H2ax^{-/-}$, $Mdc1^{-/-}$ and $Atm^{-/-}$ mice

(Supplementary Figs 1 and 2)[14,15]. $Atm^{-/-}$ thymocytes express low levels of T-cell antigen receptor (TCR)-αβ, and we found a comparable defect in $53BP1^{-/-}$ but not in $H2ax^{-/-}$ or $Mdc1^{-/-}$ thymocytes (Supplementary Figs 1c and 2b). In addition, there was a reduction in the number of γδ T cells in the $53BP1^{-/-}$ thymus and spleen (Supplementary Fig. 1d), and a marked increase in apoptosis (Supplementary Fig. 1e). Transgenic expression of TCR-αβ in the absence of endogenous recombination rescued the number of thymocytes to levels comparable to that of littermate controls (Supplementary Fig. 3), indicating that decreased cellularity of the $53BP1^{-/-}$ thymus was due to defective V(D)J recombination.

To ascertain whether recombination defects exist in $53BP1^{-/-}$, $H2ax^{-/-}$ or $Mdc1^{-/-}$ thymocytes, we initially assayed for Tcra locus integrity by three-dimensional interphase DNA-fluorescent in situ hybridization (FISH) with 5′ and 3′ bacterial artificial chromosome (BAC) probes (Fig. 1a)[16]. The probes were detected as two pairs of signals in most (>94%) wild-type, $H2ax^{-/-}$ and $Mdc1^{-/-}$ thymocytes (Fig. 1a and Supplementary Fig. 2c). More rarely, wild-type, $H2ax^{-/-}$ and $Mdc1^{-/-}$ thymocytes had lost either Cα (2 TCRVα, 1 TCRCα signal; up to 0.1% of thymocytes) or both Vα and Cα from one allele (1 TCRVα, 1 TCRCα; up to 0.7% of thymocytes). In contrast to $H2ax^{-/-}$ and $Mdc1^{-/-}$ thymocytes, $53BP1^{-/-}$ thymocytes had a sevenfold increase in the number of aberrant cells (Fig. 1a and Supplementary Fig. 2c). Even when one TCR locus was entirely lost (1 TCRVα, 1 TCRCα), simultaneous hybridization with a centromeric chromosome 14 BAC probe confirmed that both chromosomes were present (data not shown), suggesting extensive but not complete chromosome 14 degradation (see later). Thus, 53BP1 is required for the stability of the Tcra locus, although the frequency of aberrant cells is higher in $Atm^{-/-}$ mice (Fig. 1a)[16].

To determine whether extensive thymocyte cell death or the stability of the Tcra locus is affected by pro-survival factors, we bred 53BP1-deficient mice with p53-deficient ($p53^{-/-}$) or mutant Nbs1 mice lacking the pro-apoptotic carboxy-terminal portion of ($Nbs1^{tr735}$)[17]. Although these crosses reduced the number of TUNEL-positive thymocytes (Supplementary Fig. 1e and data not shown), they failed to rescue either the cellular or the developmental defects associated with 53BP1 deficiency (Supplementary Fig. 1a–d). However, further loss of the Tcra locus to levels approaching Atm deficiency was found in double-mutant $53BP1^{-/-}p53^{-/-}$ and $53BP1^{-/-}Nbs1^{tr735}$ mice (Fig. 1a). Thus, although enhanced survival is insufficient to rescue lymphocyte development in $53BP1^{-/-}$ mice (Supplementary Fig. 1), some of the $53BP1^{-/-}$ thymocytes that harbour unrepaired breaks in the TCR locus eventually die by a p53- and NBS1-sensitive mechanism.

[1]Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-1360, USA. [2]Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [3]Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [4]Laboratory of Molecular Immunology, The Rockefeller University and Howard Hughes Medical Institute, New York, New York 10021, USA. *These authors contributed equally to this work.
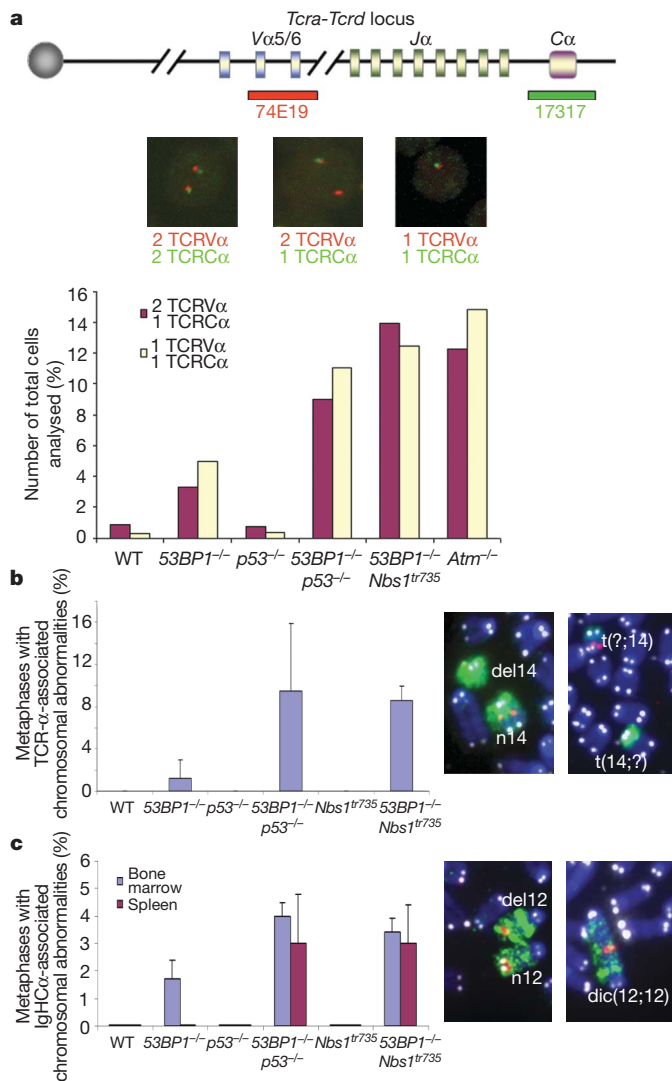
a

Figure 1 | Antigen-receptor-associated aberrations in 53BP1[−/−] lymphocytes.

a, Top panel: schematic of the *Tcra-Tcrd* locus, with positions of the BACs used for generation of DNA FISH probes indicated. Middle panel: representative examples of projections of confocal sections, analysed by three-dimensional FISH on freshly isolated thymocytes. Bottom panel: frequency at which the TCRCα or the (TCRVα + TCRCα) signal is lost from one allele in wild-type (WT), *53BP1*[−/−], *p53*[−/−], *53BP1*[−/−] *p53*[−/−], *53BP1*[−/−]*Nbs1*[tr735] and *Atm*[−/−] thymocytes (>200 cells analysed per genotype in each of two experiments). **b**, TCR-α-associated chromosomal aberrations in lymph node T cells, determined by FISH using a *Tcra*-locus-spanning BAC (red signal, right panels), a chromosome 14 paint (green signal, right panels) and a telomere-repeat-specific probe (white signal, right panels). del14, deletion of chromosome 14; dic(12;12) dicentric chromosome 12; n14, normal chromosome 14; t(?;14), translocation between chromosome ? and 14. T cells were stimulated with anti-TCR/CD28 antibodies for 48 h (error bars, s.d.; *n* ≥ 3). **c**, Frequency of IgH-associated abnormalities in metaphase spreads from bone marrow and splenic B cells from wild-type, *53BP1*[−/−], *53BP1*[−/−]*p53*[−/−] and *53BP1*[−/−]*Nbs1*[tr735] mice (error bars, s.d.; *n* ≥ 3). B220[+] bone marrow cells were cultured on irradiated S17 stromal cells in the presence of interleukin 7 for 5 days. Splenic CD43-negative B cells were cultured for 1 day with RP105, which induces proliferation but not CSR[8]. FISH was performed using probes specific for the *IghCα* locus (red signal, right panels), chromosome 12 (green signal, right panels) and telomeric repeats (white signal, right panels).

Our finding that loss of *Tcra* locus integrity is exacerbated in *53BP1*[−/−] thymocytes that are also *p53*- or *Nbs1*[tr735]-deficient led us to investigate whether such cells would maintain V(D)J-recombination-induced chromosome aberrations in mature peripheral lymphocytes as previously shown in *Atm*[−/−] mice[8]. Whereas TCR-

α-associated chromosome breaks and translocations were undetectable or rare in metaphases from wild-type and *53BP1*[−/−] lymph node T cells, they were found in *53BP1*[−/−] *p53*[−/−] and *53BP1*[−/−]*Nbs1*[tr735] doubly deficient mice at similar levels to that seen in *Atm*[−/−] mice (Fig. 1b, Supplementary Table 1 and ref. 8). Likewise, concomitant loss of *p53* or the *Nbs1* C terminus resulted in increased immunoglobulin (Ig)H-specific aberrations in *53BP1*[−/−] bone marrow, and these aberrations were maintained in the peripheral B-cell compartment (Fig. 1c and Supplementary Table 1). We conclude that the long-term persistence of V(D)J-recombination-induced aberrations in *53BP1* knockout mice is limited by an intact apoptotic checkpoint.

Although antigen-receptor-associated DSBs occur in *53BP1*[−/−] lymphocytes (Fig. 1b, c), previous semiquantitative analysis of TCR-α rearrangements showed no defect[4]. However, measurement of recombination efficiency at this locus is complicated by the fact that there are multiple gene segments (100 Vα and 61 Jα) and that TCR-α alleles undergo primary and secondary rearrangements. To examine V(D)J recombination quantitatively, we evaluated rearrangements in mice homozygous for the *Tcra*[SJ] allele, in which the 61-gene Jα cluster was replaced by two Jα segments[18] (Fig. 2a). Southern blot analysis of *Tcra*[SJ/SJ]WT (where WT represent wild type), *Tcra*[SJ/SJ] *53BP1*[−/−] and *Tcra*[SJ/SJ] *Atm*[−/−] thymocyte DNA was used to assess aberrant Jα56 (where 56 is the number of the Jα gene segment) coding end accumulation. As previously documented, Jα56 coding ends accumulate as a discrete fragment in *Tcra*[SJ/SJ] *Atm*[−/−] but not in *Tcra*[SJ/SJ]WT thymocytes, suggesting an end-joining defect in the absence of ATM[18] (Fig. 2a). Notably, multiple independent preparations of *Tcra*[SJ/SJ] *53BP1*[−/−] thymocytes displayed abnormal joining and coding end degradation, as evidenced by the Cα probe hybridizing with diffuse DNA fragments that migrated below the expected size of the Jα56 coding end (Fig. 2a). Furthermore, deoxynucleotidyl transferase (TdT) labelling of DNA ends revealed accumulation of the Jα coding ends and a broad smear of DNA that migrated below Jα61 in *Tcra*[SJ/SJ] *53BP1*[−/−] thymocytes (Fig. 2b, c). Taken together, these results suggest that in the absence of 53BP1 there is a repair defect and extensive nucleolytic processing of unrepaired V(D)J-recombination-induced coding ends.

To determine whether aberrant V(D)J recombination products were present in *53BP1* knockout thymocytes, we cloned and sequenced TCR-β, TCR-α and TCR-δ coding junctions (*n* = 273) from *53BP1*[−/−] thymocytes. We found no statistically significant difference in the overall number of deletions or insertions at TCR loci (Supplementary Table 2). Furthermore, unlike classical non-homologous end joining (NHEJ)-deficient mice, in which most joins contain short stretches of homology indicative of an alternative end-joining mechanism[19], there was no bias in micro-homology joins in *53BP1*[−/−] mice (Supplementary Table 2). Nevertheless, four of the *53BP1*[−/−] junctions contained deletions >300 base pairs (bp), four contained inserts >170 bp, and two of these junctions with insertions also showed significant degradation of V and J segments (Fig. 2d). In contrast, none of the 201 wild-type junctions analysed showed such abnormalities. Interestingly, the captured insertions in *53BP1*[−/−] junctions were not random, but were derived from intervening DNA that is normally deleted (Fig. 3d). This reintegration of excised episomes during V(D)J recombination constitutes a source of genome instability, and resembles the abnormalities reported for CSR junctions in *53BP1*[−/−] B cells[12].

Defective CSR in the absence of 53BP1 is associated with an increase in short-range intra-switch recombination[12]. To determine whether 53BP1 facilitates joining of distal gene segments during V(D)J recombination, we performed quantitative polymerase chain reaction (PCR) assays of partial (Dδ2−Jδ1 and Dδ1−Dδ2) and complete (Vδ5−Dδ1 and Vδ4−Dδ1) rearrangements. We found that short-range (Dδ2−Jδ1 and Dδ1−Dδ2) rearrangements were similar to or even more abundant in *53BP1*[−/−] than in wild-type thymocytes (Fig. 3a). In contrast, complete Vδ-to-DδJδ recombination was reduced approximately 2.5-fold in *53BP1* knockout thymocytes
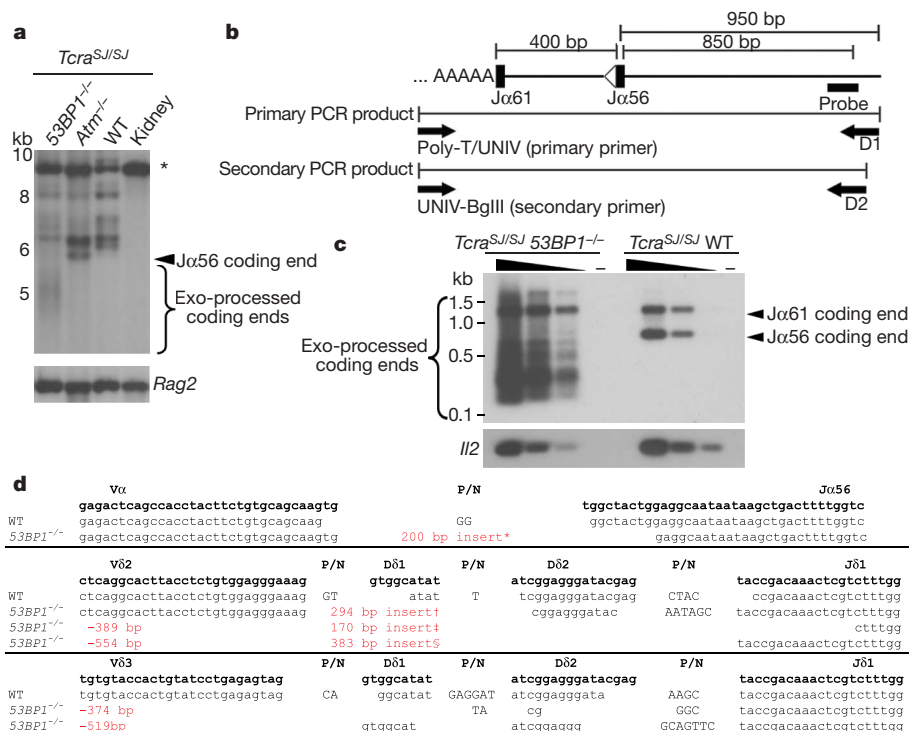
Figure 2 | Processing of Jα coding ends in 53BP1-deficient thymocytes.
a, Southern blot analysis of total thymocyte and kidney DNA digested with StuI and hybridized with a TCRCα probe. The same blot was stripped and re-probed with a *Rag2* probe as a loading control. The fragments corresponding to the germline *Tcra^SJ* allele (asterisk), the Jα56 coding ends and the molecular mass markers are indicated. **b**, Strategy for PCR amplification of free Jα coding ends captured by TdT-mediated polyadenylation. Sequences for UNIV, D1 and D2 primers are provided in Supplementary Table 4.

c, Southern blot analysis of coding ends from *Tcra^SJ/SJ* WT and *Tcra^SJ/SJ* 53BP1^−/−^ thymocytes. Serial fivefold dilutions and a mock-polyadenylated control are shown. *Il2* PCR (serial five-fold dilutions) is show as a loading control. **d**, Examples of sequences of coding joints formed by V(D)J recombination. Germline coding sequences (bold lowercase letters) are indicated at the top. Nucleotide insertions (P/N) are indicated by capital letters. Sequences of inserts are listed in Supplementary Table 2.
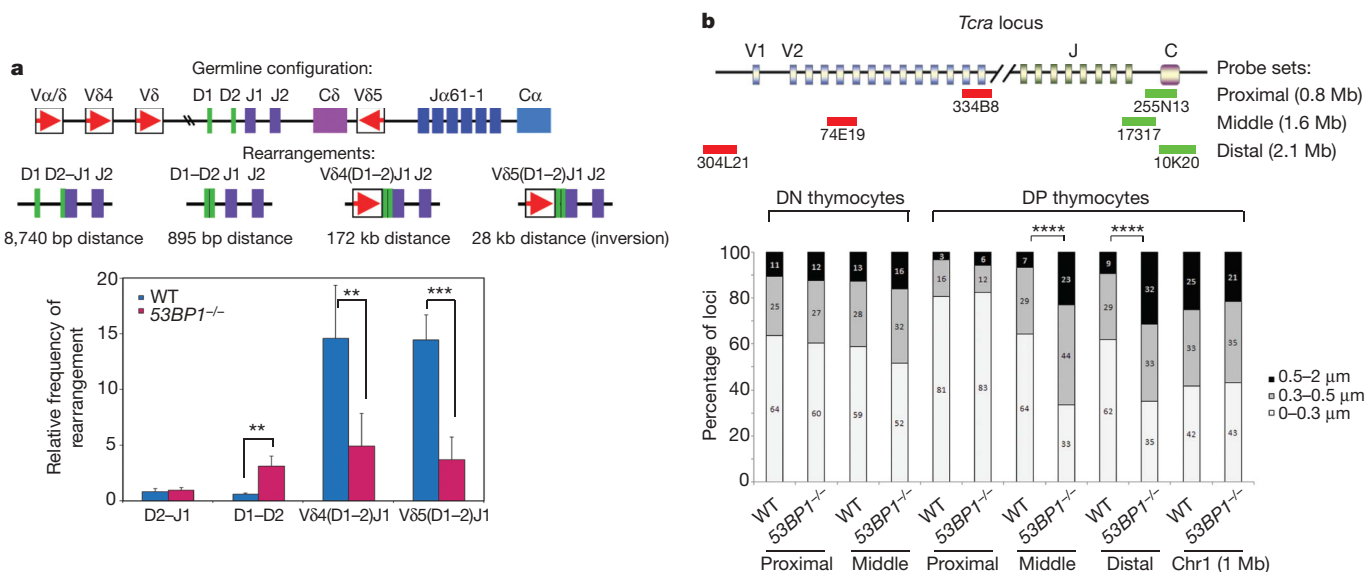

Figure 3 | Decreased efficiency of long-range V(D)J recombination and *Tcra* locus contraction in *53BP1^−/−^* thymocytes. **a**, A schematic of the mouse *Tcra-Tcrd* locus is shown in the top panel, with the four different *Tcrd* rearrangements depicted individually below. The lower panel shows the frequency of *Tcrd* locus rearrangements in total thymocytes from *53BP1^−/−^* and wild-type littermates. Quantitative assessment of genomic DNA rearrangements of Dδ1 to Dδ2, Dδ2 to Jδ1, and Vδ4 and Vδ5 to (D)Jδ1 genes were performed by qPCR and normalized to the signal of the non-rearranging DNA 3' of Jδ2. Results are averaged from four *53BP1^−/−^* and two wild-type mice with duplicate measurements; error bars, s.e.m. (wild

type versus *53BP1^−/−^*: D2–J1, $P = 0.851$; D1–D2, $P = 0.001$; V4–DJ1, $P = 0.003$; V5–DJ1, $P < 0.001$). **$P < 0.01$; ***$P < 0.001$. **b**, Distances separating the *TcrVα* and *TcrCα* loci in freshly isolated CD4/CD8 double-negative (DN) or double-positive (DP) thymocytes. The data were obtained by three-dimensional DNA FISH using the TCR-α BAC probes as indicated (top panel). As control, the probe set RP23-309A8 and RP24-336F10 (1 Mb distance) on mouse chromosome 1D was used. Light grey bars, 0–0.3 μm; dark grey bars, 0.3–0.5 μm; and black bars, 0.5–2 μm. Distance distribution histograms are shown in Supplementary Fig. 4, and Supplementary Table 3 lists sample sizes and average distances (error bars, s.d.; *****$P < 0.0001$).
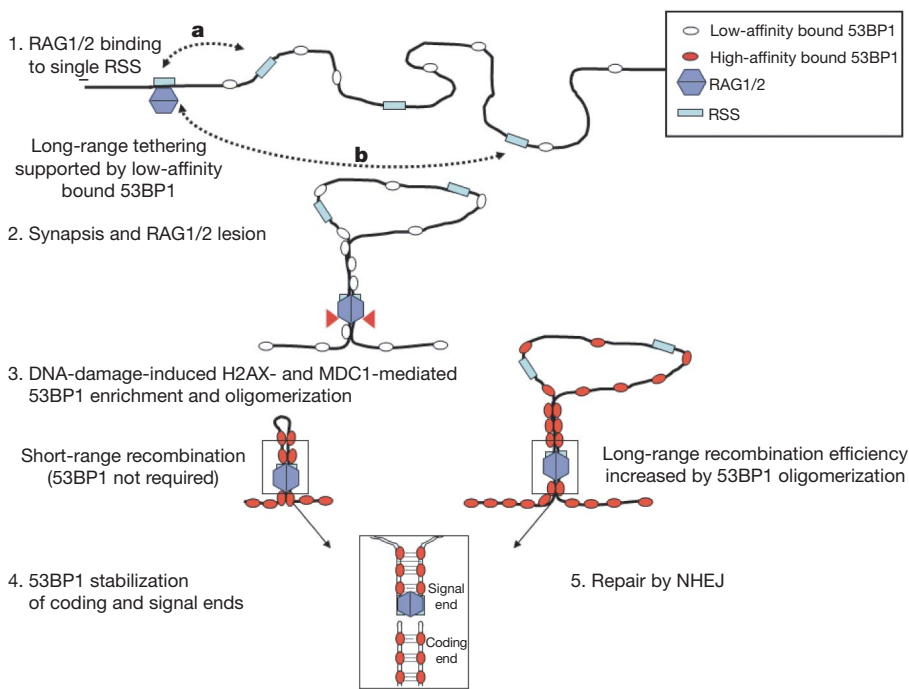
Figure 4 | Model for the role of 53BP1 in promoting and/or maintaining synapsis during V(D)J recombination. Before recombination, 53BP1 is loosely associated with chromatin (step 1). RSSs that are at a close physical distance (**a**) spend more time near the RAG1/2-bound RSS than distant RSSs (**b**), and have a high probability of synapsis whether or not 53BP1 is present. 53BP1 homo-oligomerization increases the effective local concentration of distant RSSs, thereby promoting their pairing. When RAG1/2 associates with and cleaves a single RSS, 53BP1 accumulates at the lesion and binds tightly to flanking chromatin in a manner dependent on H2AX and MDC1 (steps 2 and 3). After RAG1/2-mediated DSB formation (step 4), 53BP1 stabilizes the post-cleavage complex, thereby increasing the efficiency of NHEJ (step 5).

(Fig. 3a). To substantiate these findings further, we compared Jα usage in $53BP1^{-/-}$ and wild-type thymocytes. Long-range primary Vα–Jα rearrangements (involving at least 200 kilobases (kb)) use Jα segments at the 5′ end of the Jα array and these are replaced by shorter range secondary rearrangements that use progressively more 3′ segments at the end of the Jα locus. Comparison of Jα usage in $53BP1^{-/-}$ and wild-type thymocytes revealed that the 5′ proximal Jα segments were under-represented in $53BP1^{-/-}$ mice by at least twofold whereas distal ones were relatively increased (Supplementary Fig. 4). Taken together, these data suggest that long-range V(D)J recombination is impaired in $53BP1$ knockout mice whereas short-range recombination is not affected and may be favoured.

If distal joining was preferentially impaired, we would expect that the distances separating the TCR Vα and Jα segments would be greater in the absence of 53BP1. To study the compaction state of the *Tcra* locus, we measured distances between 5′ Vα and 3′ Cα probes in $CD4^+CD8^+$ double-positive thymocytes that are actively undergoing TCR-α recombination (Fig. 3b). Three different 5′ and 3′ BAC probe sets separated by 0.820 Mb (proximal), 1.6 Mb (middle) and 2.1 Mb (distal) were used to map spatial distances along the *Tcra* locus (Fig. 3b). The average distance (D) between proximal probes (334B8–255N13) was similar in wild-type thymocytes ($D = 0.216\,\mu M$) and $53BP1^{-/-}$ double-positive thymocytes ($D = 0.225\,\mu M$; wild-type versus $53BP1^{-/-}$, $P = 0.57$, two-tailed Wilcoxon rank sum test), and 81–83% of the alleles were separated by distances of less than 0.3 μM in both cases (Fig. 3b, Supplementary Table 3 and Supplementary Fig. 5). In contrast, the distribution of distances among intact TCR-α alleles in $53BP1^{-/-}$ double-positive thymocytes was substantially shifted towards larger separations when assessed using either middle (74E19–17317) or distal (304L21–10K20) probe sets (Fig. 3b, Supplementary Table 3 and Supplementary Fig. 5). For example, the average distance between the 5′ and 3′ ends of the *Tcra* locus was 1.4-fold greater in $53BP1^{-/-}$ compared to wild-type double-positive thymocytes (Fig. 3b; middle probes: wild type, $D = 0.273\,\mu M$, and $53BP1^{-/-}$, $D = 0.396\,\mu M$, $P < 0.0001$; distal probes: wild type, $D = 0.287\,\mu M$, and $53BP1^{-/-}$, $D = 0.426\,\mu M$, $P < 0.0001$). In $53BP1^{-/-}$ double-positive thymocytes, middle and distal genes were separated by 0.5–2 μM in 23–32% of nuclei compared to only 7–9% in wild type. In contrast to double-positive thymocytes, the distribution of spatial distances was similar for loci not undergoing recombination (chromosome band

1D and the *Tcra* locus in $CD4^{-/-}CD8^{-/-}$ double-negative thymocytes; Fig. 3b and Supplementary Table 3). We conclude that the distal portion of the *Tcra* locus is in a more extended state in $53BP1^{-/-}$ double-positive thymocytes.

RAG1/2 initiates recombination by binding to a single recombination signal sequence (RSS) and introduces a lesion in the target DNA[20,21] (Fig. 4, steps 1 and 2). It is believed that a synaptic complex is produced when a second RSS is captured[20,21]. 53BP1 interacts constitutively with chromatin in the absence of DNA damage by means of its methyl-histone-binding tudor domain[22–24]. 53BP1 oligomerization[25] may promote a basal level of RSS synapsis by increasing the effective local concentration of the RSSs, thereby promoting their pairing (Fig. 4, steps 1 and 2). 53BP1 has also been reported to interact with motor proteins[26], which may enhance chromatin mobility. On DNA damage the affinity of 53BP1 for DNA is increased by mechanisms that involve interaction with H2AX and MDC1 (refs 1, 2, 24). We speculate that the initiating lesions in V(D)J recombination promote high-affinity binding of 53BP1, and that this further amplifies 53BP1 homo-oligomerization and therefore the efficiency of long-range synapsis (Fig. 4, step 3). 53BP1 recruitment to DSB-flanking chromatin could also have an important role in the post-cleavage phase of the recombination reaction by protecting coding ends from premature release from the synaptic complex and from degradation before the breaks are repaired by NHEJ (Fig. 4, steps 4 and 5). Similarly, joining of distal switch regions during CSR would be favoured by high-affinity binding of 53BP1 and local chromatin contraction to bring together damaged DNA ends[3,4,10–12].

Such a model would be compatible with the NHEJ defect in the absence of 53BP1 owing to loss of both 'basal' and DNA-damage-induced RSS tethering, and with the mild joining defect in the absence of H2AX or MDC1, where only the DSB-induced amplification of the affinity of 53BP1 for damaged chromatin is impaired. We propose that 53BP1-mediated chromosome tethering and/or mobility evolved to facilitate chromosomal end joining, and thereby favour long-range interactions between DNA ends that have an otherwise low probability of encountering each other[27].

METHODS SUMMARY

Mice. $p53^{-/-}$ and $Rag2^{-/-}$ mice were purchased from Taconic Laboratories and TCR-transgenic mice expressing a receptor specific for a fragment of pigeon cyto-

chrome *c TCR(AND)* were obtained from Jackson Laboratories. *53BP1*$^{-/-}$ and *Mdc1*$^{-/-}$ mice were obtained from J. Chen and *Atm*$^{-/-}$ mice from A. Wynshaw-Boris; and *Nbs1*^{tr735} and *H2ax*$^{-/-}$ mice have been described previously[14,17].

Metaphase and interphase FISH. Metaphase spreads from lymphocytes were prepared and analysed by FISH as previously described[8]. Three-dimensional FISH on thymocytes using *TcrVα*- and *TcrCα*-locus-specific probes was performed as described[8]. Confocal image stacks were collected (Zeiss LSM510 META) and the three-dimensional distance between red and green spots was measured using customized software (MIPAV, CIT/NIH).

Flow cytometry. Thymocytes were washed twice in HBSS containing 0.1% BSA and 0.1% NaN$_3$, and stained with antibodies specific for CD4, CD8, TCR-β (H57-597) and TCR-γδ (GL3; BD PharMingen).

PCR and junction analysis. Genomic DNA was isolated from thymocytes. DNA (50 ng) was PCR amplified with a combination of gene-specific primers (Supplementary Table 4) using the Power Sybr Green Master Mix kit (Applied Biosystems) for quantitative PCR (qPCR). Real-time qPCR was performed in duplicate and data were collected on an ABI 9700 sequence analyser and analysed using the SDS 2.2 software (Applied Biosystems). For sequence analysis, TCR coding junctions were amplified using the indicated primers (Supplementary Table 4 and previously published[28]) and cloned using a TA cloning kit (Invitrogen). TCRJα usage was determined as described previously[29].

Southern blot analysis. Southern blot analysis was carried out as described previously[18].

TdT-mediated polyadenylation and PCR amplification of free Jα coding ends. Genomic DNA from thymocytes was treated with TdT (New England Biolabs), amplified by PCR and analysed by Southern blotting.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Stucki, M. & Jackson, S. P. γH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair* 5, 534–543 (2006).
2. Fernandez-Capetillo, O., Lee, A., Nussenzweig, M. & Nussenzweig, A. H2AX: the histone guardian of the genome. *DNA Repair* 3, 959–967 (2004).
3. Manis, J. P. et al. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nature Immunol.* 5, 481–487 (2004).
4. Ward, I. M. et al. 53BP1 is required for class switch recombination. *J. Cell Biol.* 165, 459–464 (2004).
5. Petersen, S. et al. AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching. *Nature* 414, 660–665 (2001).
6. Chen, H. T. et al. Response to RAG-mediated VDJ cleavage by NBS1 and γ-H2AX. *Science* 290, 1962–1965 (2000).
7. Reina-San-Martin, B., Chen, H. T., Nussenzweig, A. & Nussenzweig, M. C. ATM is required for efficient recombination between immunoglobulin switch regions. *J. Exp. Med.* 200, 1103–1110 (2004).
8. Callen, E. et al. ATM prevents the persistence and propagation of chromosome breaks in lymphocytes. *Cell* 130, 63–75 (2007).
9. Reina-San-Martin, B. et al. H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.* 197, 1767–1778 (2003).
10. Ramiro, A. R. et al. Role of genomic instability and p53 in AID-induced *c-myc–Igh* translocations. *Nature* 440, 105–109 (2006).
11. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* 21, 201–214 (2006).
12. Reina-San-Martin, B., Chen, J., Nussenzweig, A. & Nussenzweig, M. C. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1$^{-/-}$ B cells. *Eur. J. Immunol.* 37, 235–239 (2007).
13. Ward, I. M., Minn, K., van Deursen, J. & Chen, J. p53 binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. *Mol. Cell. Biol.* 23, 2556–2563 (2003).
14. Celeste, A. et al. Genomic instability in mice lacking histone H2AX. *Science* 296, 922–927 (2002).
15. Lou, Z. et al. MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* 21, 187–200 (2006).
16. Matei, I. R. et al. ATM deficiency disrupts *Tcra* locus integrity and the maturation of CD4$^+$CD8$^+$ thymocytes. *Blood* 109, 1887–1896 (2007).
17. Difilippantonio, S. et al. Distinct domains in Nbs1 regulate irradiation-induced checkpoints and apoptosis. *J. Exp. Med.* 204, 1003–1011 (2007).
18. Huang, C. Y. et al. Defects in coding joint formation *in vivo* in developing ATM-deficient B and T lymphocytes. *J. Exp. Med.* 204, 1371–1381 (2007).
19. Gu, Y. et al. Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity* 7, 653–665 (1997).
20. Curry, J. D., Geier, J. K. & Schlissel, M. S. Single-strand recombination signal sequence nicks *in vivo*: evidence for a capture model of synapsis. *Nature Immunol.* 6, 1272–1279 (2005).
21. Jones, J. M. & Gellert, M. Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *EMBO J.* 21, 4162–4171 (2002).
22. Huyen, Y. et al. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* 432, 406–411 (2004).
23. Botuyan, M. V. et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127, 1361–1373 (2006).
24. Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *J. Cell Biol.* 170, 201–211 (2005).
25. Adams, M. M. et al. 53BP1 oligomerization is independent of its methylation by PRMT1. *Cell Cycle* 4, 1854–1861 (2005).
26. Lo, K. W. et al. The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation. *J. Biol. Chem.* 280, 8172–8179 (2005).
27. Jhunjhunwala, S. et al. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* 133, 265–279 (2008).
28. Talukder, S. R., Dudley, D. D., Alt, F. W., Takahama, Y. & Akamatsu, Y. Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *Nucleic Acids Res.* 32, 4539–4549 (2004).
29. Huang, C. Y. & Sleckman, B. P. Developmental stage-specific regulation of TCR-α-chain gene assembly by intrinsic features of the TEA promoter. *J. Immunol.* 179, 449–454 (2007).

METHODS

Metaphase and interphase FISH. Metaphase spreads from cultured lymphocytes were prepared and analysed by FISH[8] using BAC probes containing the *Tcra* locus (TCRCα, 17317 and TCRVα, RP24-74E19), the *Igh* locus (IgHVα, 224M14), IgH Cα (from Cγ1 to 3′of Cα) and a telomere-repeat-specific peptide nucleic acid probe (Applied Biosystems). Three-dimensional FISH on freshly isolated thymocytes using *TcrVα*- and *TcrCα*-locus-specific probes was performed as described[8]. Confocal image stacks were collected using a Zeiss LSM510 META microscope equipped with a ×63 Plan-Apochromat (NA of 1.4) objective lens with 0.07 μm *x*–*y* pixel sampling, an optical slice thickness of 0.8 μm, and a *Z*-step size of 0.2 μm. Volumes of interest were drawn around FISH spots to be measured. Within the individual volumes of interest, the centre of intensity mass was calculated for each FISH spot and the three-dimensional distance between red and green spots was measured using customized software (MIPAV, CIT/NIH). Tetraspeck fluorescent microspheres (Molecular Probes/Invitrogen) 0.1 μm and 0.5 μm in diameter were imaged using the same microscope parameters and used to model the point spread function of the microscope and the fluorescence channel alignment, and to determine the validity of the MIPAV software to accurately measure the centre of intensity mass of a fluorescence object. The minimum measurable distance between the two fluorescence points was 70 nm in the lateral dimension. For distance measurements the following BAC probe sets were used: RP24-334B8 and RP23-255N13 (proximal), RP24-74E19 and 17317 (middle), RP23-304L21 and RP23-10K20 (distal), and as a control RP23-309A8 and RP24-336F10, which are separated by 1 Mb and map to mouse chromosome 1D.

Flow cytometry. Thymocytes were washed twice in HBSS containing 0.1% BSA and 0.1% NaN₃ and stained with antibodies specific for CD4, CD8, TCR-β (H57-597) and TCR-γδ (GL3; BD PharMingen).

PCR and junction analysis. Genomic DNA was isolated from thymocytes using a Qiagen DNA isolation kit (Qiagen). DNA (50 ng) was PCR amplified with a combination of gene-specific primers (Supplementary Table 3) using Platinum Taq polymerase (Invitrogen) for the standard, or the Power Sybr Green Master Mix kit (Applied Biosystems) for qPCR. Real-time qPCR was performed in duplicate and data were collected on an ABI 9700 sequence analyser and analysed using the SDS 2.2 software (Applied Biosystems). For each assay, aliquots of DNA were analysed for a control, non-rearranging DNA 3′ of Jβ2. The cycle threshold numbers for each primer combinations (C_t^{exp}) and for the control amplification (C_t^{ctr}) were used to calculate the absolute amount of PCR signal. The relative ratios of each rearrangement were averaged and plotted together with the standard error of the mean. For sequence analysis, 150 ng thymocyte DNA was used to amplify coding joints, which were cloned using a TA cloning kit (Invitrogen). Primers for TCR-β coding joints were published[28]. To PCR amplify TCR V–DJδ coding joint sequences, Vδ primers were used as forward primers with the primer 3Jδ1 as the reverse primer (Supplementary Table 4). Junctions from *Tcra*^SJ mice were amplified using the indicated primers (Supplementary Table 4). TCRJα usage was determined as described previously[29].

Southern blot analysis. Southern blot analysis was carried out as described previously[18]. In brief, 10 μg total thymus genomic DNA was digested with StuI (New England Biolabs). The digest products were run out on a 1% agarose TAE gel, transferred onto Zeta-Probe GT membrane and probed with the TCRCα for coding joints and coding ends. The same membrane was stripped and re-probed with RAG2 as a loading control.

TdT-mediated polyadenylation and PCR amplification of free Jα coding ends. One microgram of genomic DNA was treated with TdT (New England Biolabs) according to the manufacturer's protocol at a final concentration of 5 μM dATP. Terminal deoxynucleotidyl transferase was subsequently terminated by heating to 70 °C for 15 min. Two per cent of the total volume of each polyadenylation reaction was then used for primary amplification using D1 and T17-UNIV primers. Conditions were 95 °C for 5 min followed by 15 cycles of 95 °C for 1 min, 57 °C for 45 s and 72 °C for 45 s, and then a final 5 min extension step at 72 °C. Two per cent by volume of each primary amplification reaction was then serially diluted fivefold in water. One microlitre of the original primary reaction and of each serial dilution were used as a template for a secondary amplification step using primers D2 and UNIV-BglII. Reaction conditions were the same as for the primary reactions, but the total number of cycles was increased to 30 cycles total. Fifteen microlitres of each secondary reaction was run out on a 1% agarose TBE gel and then transferred onto Zeta-Probe GT membrane (Bio-Rad). The membrane was subsequently hybridized with ³²P-labelled DS850 oligonucleotide. For IL-2 loading control PCR reactions, aliquots of the DNA samples used for polyadenylation were serially diluted fivefold and amplified as described previously[30]. Ten microlitres of each PCR product was run out on a 1% agarose TBE gel and then transferred overnight onto a Zeta-Probe GT membrane (Bio-Rad). The oligonucleotides D1, D2, UNIV-BglII, T17-UNIV and DS850 are listed in Supplementary Table 4.

TUNEL staining. Thymi were fixed in buffered 10% formalin, and paraffin sections were stained with haematoxylin and eosin. Apoptotic nuclei were detected with TdT-mediated dUTP-biotin nick labelling.

30. Bredemeyer, A. L. *et al.* ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466–470 (2006).

LETTERS

The role of HLA-DQ8 β57 polymorphism in the anti-gluten T-cell response in coeliac disease

Zaruhi Hovhannisyan[1], Angela Weiss[2], Alexandra Martin[1], Martina Wiesner[3], Stig Tollefsen[4], Kenji Yoshida[5], Cezary Ciszewski[1], Shane A. Curran[1], Joseph A. Murray[6], Chella S. David[6], Ludvig M. Sollid[4,7], Frits Koning[3], Luc Teyton[5] & Bana Jabri[1]

Major histocompatibility complex (MHC) class II alleles HLA-DQ8 and the mouse homologue I-A^{g7} lacking a canonical aspartic acid residue at position β57 are associated with coeliac disease[1,2] and type I diabetes[3,4]. However, the role of this single polymorphism in disease initiation and progression remains poorly understood. The lack of Asp 57 creates a positively charged P9 pocket, which confers a preference for negatively charged peptides. Gluten lacks such peptides, but tissue transglutaminase (TG2) introduces negatively charged residues at defined positions into gluten T-cell epitopes by deamidating specific glutamine residues[5,6] on the basis of their spacing to proline residues[7]. The commonly accepted model, proposing that HLA-DQ8 simply favours binding of negatively charged peptides, does not take into account the fact that TG2 requires inflammation for activation[8] and that T-cell responses against native gluten peptides are found[9,10], particularly in children[11]. Here we show that β57 polymorphism promotes the recruitment of T-cell receptors bearing a negative signature charge in the complementary determining region 3β (CDR3β) during the response against native gluten peptides presented by HLA-DQ8 in coeliac disease. These T cells showed a crossreactive and heteroclitic (stronger) response to deamidated gluten peptides. Furthermore, gluten peptide deamidation extended the T-cell-receptor repertoire by relieving the requirement for a charged residue in CDR3β. Thus, the lack of a negative charge at position β57 in MHC class II was met by negatively charged residues in the T-cell receptor or in the peptide, the combination of which might explain the role of HLA-DQ8 in amplifying the T-cell response against dietary gluten.

A specific assessment of the role of MHC class II alleles in antigen-specific T-cell responses is possible in coeliac disease, because it is caused by a single and well-defined antigen, gluten, which is present in wheat, rye and barley (reviewed in ref. 12). Gluten is rich in glutamine and proline, and the high content of proline renders gluten resistant to intestinal enzymatic digestion, explaining the persistence of intact peptides for immune recognition. Notably, despite the presence of few negatively charged amino acid residues in gluten and the expression of other MHC alleles potentially capable of presenting gluten peptides, all gluten-specific CD4$^+$ T cells isolated from the intestinal mucosa of adult patients are restricted by HLA-DQ2 or HLA-DQ8 (refs 10, 13 and 14), and most recognize a dominant peptide with a deamidated Gln—that is, a negatively charged Glu residue[5,6].

To study the T-cell response to native and deamidated gluten peptides (Fig. 1a), we immunized humanized HLA-DQ8 mice with the native 24-amino-acid gluten peptide 219–242 from the α2 (AJ133612) gliadin comprising the commonly recognized DQ8-α–I epitope (QGSFQPSQQ)[10], or with a peptide deamidated at positions 229 and 237, the known targets of TG2 (ref. 6). As previously reported[15], the deamidated peptide failed to induce superior effector or recall proliferative (Fig. 1b, left panel and Supplementary Fig. 1a, left panel) or cytokine (Supplementary Fig. 1b, left panels) CD4$^+$ T-cell responses. Similar responses were also observed for lower, limiting doses of immunizing peptide (data not shown). However, under conditions of competition, when the two peptides were administered simultaneously, we found that the response against the deamidated peptide largely predominated over the native peptide (Fig. 1b, right panel and Supplementary Fig. 1, right panels). When peptides with single Gln to Glu substitutions at position 229 or 237 (Fig. 1a) were compared in the same competition assay against the native peptide, the E237 peptide behaved like the deamidated peptide, whereas the E229 peptide resembled the native peptide (Fig. 1b middle and lower panels). In accordance with previous studies concluding that the Glu residue at position 229 binds to pocket 1 (P1) of HLA-DQ8, whereas Glu at position 237 binds to pocket 9 (P9)[6,9,16], these results indicate that an acidic residue at position P9 confers a functional advantage with respect to immunization.

Using a panel of T-cell hybridomas derived from immunized lymph nodes, we compared the specificity and T-cell receptor (TCR) repertoire of CD4$^+$ T cells elicited by the native and the deamidated peptides. Immunization with native and deamidated peptides induced crossreactive T cells, in addition to T cells recognizing specifically one or the other form of the peptide. Notably, nearly a quarter of the hybridomas generated against the native peptide showed a 'heteroclitic' response[17] against the deamidated peptide, in that they responded much better to the deamidated than to the native peptide, with a shift in the dose response curve of up to 100-fold (Fig. 1c and Supplementary Fig. 2). Immunization with the native peptide, however, did not induce T cells that recognized the deamidated peptide exclusively, confirming the absence of in vivo deamidation in this experimental setting. Importantly, a heteroclitic response was never observed after immunization with the deamidated peptide.

Analysis of the TCR repertoire (Fig. 1d) revealed that most T cells recognizing the native peptide—whether elicited by immunization with the native or the deamidated peptides—preferentially used a limited set of Vβ chains including Vβ11, Vβ5, Vβ14 and Vβ2 T cells. T cells that exclusively recognized the native peptide predominantly expressed Vβ11 (45%) and to a lesser degree Vβ14 (15%), whereas T cells that equivalently recognized the two peptides mostly expressed
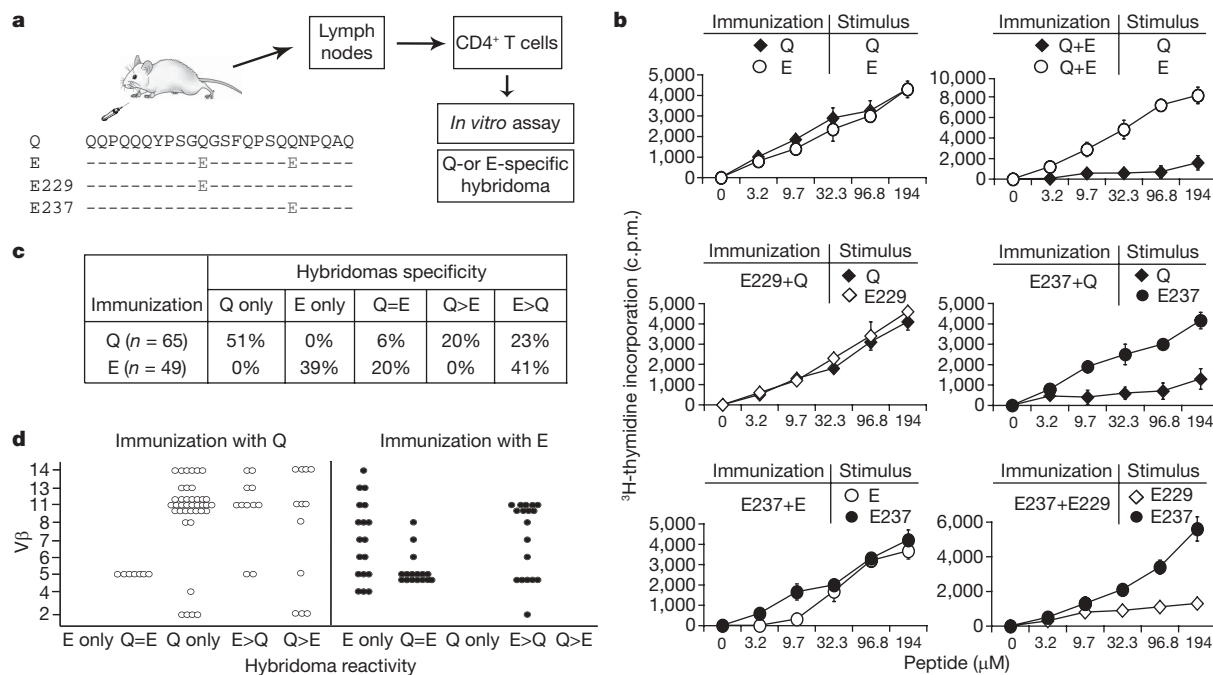
Figure 1 | Native and deamidated gluten α2-219–242 peptides recruit distinct, yet overlapping TCR repertoires. a, Humanized HLA-DQ8 transgenic mice were immunized with native (Q) and deamidated (E) versions of the 24-amino-acid gluten peptide α2-219–242 and two E peptide analogues with a Glu residue either in position 229 (E229) or 237 (E237). The draining lymph nodes were collected 8 days after immunization and CD4$^+$ T cells were purified to perform functional assays and derive α2-219–242 gluten-peptide-specific HLA-DQ8-restricted T-cell hybridomas. **b**, Purified CD4$^+$ T cells were tested for proliferative *in vitro* responses with the indicated peptide at the specified concentrations in the presence of irradiated (30 Gy) spleen cells from non-immunized humanized HLA-DQ8 mice. Responses are given in c.p.m. from the mean ^{3}H-thymidine incorporation in triplicate cultures ± s.d. $n = 3$ mice per each condition of immunization. The data are representative of three independent experiments for upper panels, and two independent experiments for middle

and lower panels. **c**, A panel of α2-219–242 gluten-peptide-specific CD4$^+$ T hybridomas was derived after immunization with native (Q) or deamidated (E) peptide alone. Hybridoma reactivity was defined by measuring IL-2 secretion. Q and E hybridomas responded exclusively to native and deamidated peptides, respectively. Q = E, hybridomas responded equally well to native and deamidated peptides; Q > E, hybridomas responded 10–100 times better to native than to deamidated peptide; E > Q, hybridomas responded 10–100 times better to deamidated than to native peptide. The percentage of T-cell hybridomas obtained in each class is indicated. Importantly, assessing the quality of the immunization, no E-specific hybridomas were obtained after Q immunization, and conversely no Q-specific hybridomas were obtained after E immunization **d**, TCR β-chain variable region usage of individual hybridomas was defined by flow cytometry after immunization with Q (open circles, left panel) and E peptides (filled circles, right panel).

Vβ5 (75%). In marked contrast, T cells specific for the deamidated peptide showed a broad TCR repertoire with no preferential Vβ (Fig. 1d) or Vα chain use (data not shown).

By studying the response of T-cell hybridomas to amino- and carboxy-terminal deletion mutant peptides, we identified five different patterns. Most (82%) of the T cells specific for the native peptide responded to the minimal ^{229}QGSFQPSQQ^{237} 9-amino-acid determinant (pattern 1), whereas the remaining clones responded to the minimal 9-amino-acid ^{228}GQGSFQPSQ^{236} determinant (pattern 2; Supplementary Table 1, lower panel). Hybridomas responding equally well to native and deamidated peptides also showed a relatively restricted response pattern; that is, pattern 1 and 4 accounted for 78% of these hybridomas (Supplementary Table 1, lower panel). In contrast, T cells specific for the deamidated peptide showed at least four different response patterns, none of which was predominant (Supplementary Table 1, lower panel). Notably, response pattern 2, characterized by a response to ^{228}GEGSFQPSQ^{236}, was never observed for T cells specific for the deamidated peptide (Supplementary Table 1, lower panel), indicating a requirement for the acidic residue at position 237. The diversity of response patterns observed with T cells specific for the deamidated peptide might be due to shifts in the peptide register, as the structure of MHC class II molecules allows peptides to frameshift within the MHC groove[18]. Alternatively, a single register might be occupied by the deamidated peptide, but the diverse TCRs could contact different residues. Altogether, these results indicate that recognition of the native peptide imposes strict requirements leading to limited TCR use, whereas

by displaying several epitopes by a yet undetermined mechanism, the deamidated peptide recruits a far broader TCR repertoire.

Examination of TCR sequences revealed the notable conservation of a negative charge at position 3 of CDR3β in most T cells recognizing the native peptide, including 92% of those specific for the native peptide, and 60–92% of those that responded to both the native and the deamidated peptides (Table 1). Importantly, 91% of the negative charges were non-germline-encoded, indicating an antigen-driven selection process (Supplementary Table 2). During TCR V(D)J recombination, further diversity is generated by the stochastic addition and deletion of palindromic and non-germline nucleotides at the coding joints between the gene segments. Specifically, here more than 35% of the non-germline-encoded negatively charged residues resulted from one non-germline nucleotide addition, 19% from three palindromic nucleotide additions and the rest from two non-germline, three non-germline, or one palindromic nucleotide additions and from the combination of two palindromic and one non-germline nucleotide additions (Supplementary Table 2). Interestingly, published TCR sequences from early islet infiltrates in NOD mice are also characterized by a conserved negative charge at position 3 of the CDR3β loop[19,20]. In contrast, only 29% of T cells specific for the deamidated gluten peptide had a negative charge at that position (Table 1). A possible model accounting for these findings is that HLA-DQ8- and IA^{g7}-restricted TCRs against peptides lacking acidic residues must use a negative charge in their CDR3β to stabilize peptide–MHC (pMHC) complexes through charge–charge interactions.

Table 1 | Negative charge in the CDR3β loop of native gluten-reactive T-cell hybridomas

Hybridoma	Specificity	Vβ	Jβ	Length	CDR3β	
Q$_{R17}$ (FJ410957)	Q	3	J1–1	6	CAS Ss**d**iEV	FFGKG
Q$_{P13}$ (FJ410958)	Q	11	J2–7	6	CAS SL**d**sEQ	YFGPG
Q$_{N17}$ (FJ410959)	Q	11	J2–7	6	CAS SL**d**fEQ	YFGPG
Q$_{Q21}$ (FJ410960)	Q	11	J1–1	6	CAS SL**d**TEV	FFGKG
Q$_{O14}$ (FJ410961)	Q	11	J2–1	7	CAS SL**e**nAEQ	FFGPG
Q$_{N19}$ (FJ410962)	Q	11	J2–1	7	CAS SL**e**nAEQ	FFGPG
Q$_{R5}$ (FJ410963)	Q	11	J1–3	8	CAS SL**d**SGNTL	YFGEG
Q$_{R11}$ (FJ410967)	Q	14	J2–5	8	CAW SlpGGDTQ	YFGPG
Q$_{P6}$ (FJ410965)	Q	5.1	J2–1	9	CAS SL**e**vNYAEQ	FFGPG
Q$_{R28}$ (FJ410964)	Q	11	J2–1	9	CAS SL**EL**GnAEQ	FFGPG
Q$_{R38}$ (FJ410966)	Q	5.1	J2–4	9	CAS SL**e**TGQNTL	YFGAG
Q$_{P18}$ (FJ410969)	Q	1	J2–7	9	CAS SQ**dt**GagEQ	YFGPG
Q$_{R19x}$ (FJ410968)	Q	1	J2–7	9	CAS SQ**dt**GagEQ	YFGPG
E$_{G5}$ (FJ410970)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
Q$_{O3}$ (FJ410971)	Q=E	5.1	J1–2	9	CAS SL**e**iaNSDY	TFGSG
E$_{J3}$ (FJ410972)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
E$_{F18}$ (FJ410973)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
Q$_{N8}$ (FJ410974)	Q=E	5.1	J1–2	9	CAS SL**e**iaNSDY	TFGSG
E$_{B18}$ (FJ410975)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
E$_{D5S1}$ (FJ410976)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
Q$_{R18}$ (FJ410977)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
E$_{F15}$ (FJ410978)	Q=E	5.1	J1–2	9	CAS SL**e**TANSDY	TFGSG
Q$_{F4}$ (FJ410979)	Q=E	5.1	J1–2	9	CAS SL**e**vaNSDY	TFGSG
Q$_{Q34}$ (FJ410980)	Q=E	5.1	J2–4	9	CAS SL**e**sSQNTL	YFGAG
E$_{K1}$ (FJ410981)	Q=E	8.1	J2–5	10	CAS gd**l**TGGQDTQ	YFGPG
Q$_{C10}$ (FJ410982)	Q=E	11	J2–2	10	CAS SL**D**RgdTGQL	YFGEG
Q$_{N5}$ (FJ410983)	Q>E	2	J2–4	8	CTC SA**e**aGNTL	YFGAG
Q$_{P9}$ (FJ410984)	Q>E	14	J2–3	8	CAW SltSAETL	YFGSG
Q$_{R31}$ (FJ410985)	Q>E	14	J2–3	8	CAW SlpSAETL	YFGSG
Q$_{N13}$ (FJ410986)	Q>E	14	J1–5	8	CAW Sl**e**agQAP	LFGEG
Q$_{Q18}$ (FJ410987)	Q>E	2	J1–6	10	CTC SA**ds**GVNSPL	YFAAG
E$_{F12}$ (FJ410988)	E	8.1	J1–2	8	CAS g**D**saNSDY	TFGSG
E$_{R7}$ (FJ410989)	E	8.1	J1–2	8	CAS g**D**saNSDY	TFGSG
E$_{T18}$ (FJ410990)	E	14	J2–5	8	CAW SlpGGDTQ	YFGPG
E$_{K2}$ (FJ410991)	E	7	J2–7	9	CAS SLgGQGYEQ	YFGPG
E$_{E6}$ (FJ410992)	E	5.1	J2–5	9	CAS SL**r**gNQDTQ	YFGPG
E$_{E18}$ (FJ410993)	E	6	J2–4	9	CAS SpaGSQNTL	YFGAG
E$_{P1x}$ (FJ410994)	E	5.1	J2–7	9	CAS SlwDRdYEQ	YFGPG
E$_{T11}$ (FJ410995)	E	5.1	J2–3	9	CAS SL**e**itAETL	YFGSG
E$_{P4}$ (FJ410996)	E	6	J1–3	10	CAS SITGtSGNTL	YFGEG
E$_{A16S1}$ (FJ410997)	E	6	J2–7	10	CAS SIppqsSYEQ	YFGPG
E$_{S1}$ (FJ410998)	E	5.1	J1–2	10	CAS SlGTGk**d**SDY	TFGSG
E$_{Q10}$ (FJ410999)	E	11	J1–2	10	CAS SL**e**aGtNSDY	TFGSG
E$_{H6}$ (FJ411000)	E	11	J1–2	10	CAS SL**e**aGtNSDY	TFGSG
E$_{A39S2}$ (FJ411001)	E	2	J2–2	11	CTC SA**d**DRGNTGQL	YFGEG
Q$_{Q31-2}$ (FJ411002)	E>Q	11	J2–7	7	CAS SL**d**gYEQ	YFGPG
E$_{E9}$ (FJ411003)	E>Q	11	J2–1	8	CAS SL**e**GGAEQ	FFGPG
E$_{B4S2}$ (FJ411004)	E>Q	11	J1–6	8	CAS SL**EG**NSPL	YFAAG
E$_{C25S3}$ (FJ411005)	E>Q	14	J1–1	9	CAW SrGTtNTEV	FFGKG
Q$_{R34}$ (FJ411006)	E>Q	11	J1–5	10	CAS SL**d**GrgNQAP	LFGEG
Q$_{R15}$ (FJ411007)	E>Q	11	J2–2	11	CAS SL**dg**RsNTGQL	YFGEG
E$_{C53S3}$ (FJ411008)	E>Q	11	J2–3	12	CAS SL**e**aGGrgAETL	YFGSG
E$_{F3}$ (FJ411009)	E>Q	11	J2–3	12	CAS SL**e**aGGrgAETL	YFGSG
E$_{C42S1}$ (FJ411010)	E>Q	11	J2–3	12	CAS SLnGGgrgAETL	YFGSG

CDR3β sequences from T-cell hybridomas characterized in Fig. 1 are presented. Hybridomas names are indicated in capital letter Q when they were derived from immunization with native peptide, and in capital letter E when they were derived from immunization with deamidated peptide. GenBank accession numbers are in parentheses. Junctionally encoded amino acids are shown as lower-case. Amino acid residues at position 3 are bold when they are negatively charged.

To test whether the negative charge at position 3 of CDR3β is required for recognition of the native peptide, we generated a TCR mutant lacking the conserved negative charge (Fig. 2a), as well as a negatively charged HLA-DQ8 mutant in which an Asp was introduced at position β57 (Fig. 2b). The TCR was obtained from a representative Vβ5.1/Vα17 hybridoma that responded equivalently to the native and deamidated peptides and was retrovirally transferred

into 58α⁻β⁻ T cells. Notably, mutation of the conserved negative charge at position 3 of CDR3β (Vβ5.1 (E→Q)) resulted in complete loss of recognition of the native peptide, whereas the response to the deamidated peptide was preserved (Fig. 2a, middle panel). As a control, mutation of a non-conserved negative charge at position 8 of CDR3β (Vβ5.1 (D→N)) did not alter the response to either native or deamidated peptides (Fig. 2a, right panel). We next performed the reciprocal experiment in which HLA-DQ8 was mutated to introduce an Asp at position β57. Using T-cell hybridomas responding equally well to native and deamidated peptides, we observed a marked loss of response to the native peptide (Fig. 2b, left panel). Loss of recognition of the native peptide with the Asp 57 DQ8 mutant was also confirmed with three hybridomas recognizing selectively native peptides (data not shown). In contrast, recognition of the deamidated peptide was only moderately affected (Fig. 2b, right panel), consistent with previous studies showing that MHC alleles carrying an Asp at position β57 can bind peptides harbouring acidic residues. These results support the hypothesis that HLA-DQ8 β57 polymorphism can elicit strong responses against non-negatively charged peptides by recruiting TCRs that express a negative charge at position 3 of CDR3β.

The amplitude of the T-cell response results from the combined strength of the binding of peptide to MHC and the binding of TCR to pMHC complexes. The strong response elicited by the native peptide was surprising given the lack of a negative charge to match the positive P9 pocket in HLA-DQ8. Indeed, whereas the deamidated peptide readily competed with the binding of indicator peptide to HLA-DQ8, binding of the native peptide was below detection (Supplementary Fig. 3a). When indirectly measured by hybridoma stimulation using a clone responding to both peptides, the binding of the native peptide exhibited similar on-rate but increased off-rate (Supplementary Fig. 3b). Overall these results are in accordance with
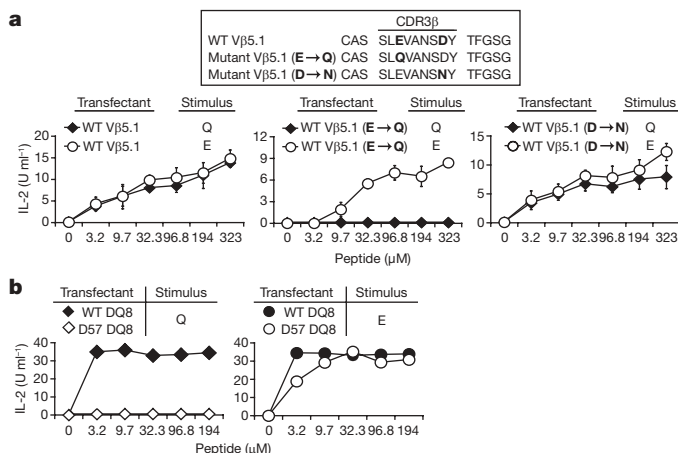

Figure 2 | Negative charge at position 3 of the TCR CDR3β loop is critical for native gluten peptide reactivity. a, Generation of TCR-Vβ5.1-chain mutants from a representative Q = E hybridoma with either substitution of Glu to Gln at position 3 (Vβ5.1 (E→Q)) or Asp to Asn at position 8 (Vβ5.1 (D→N)). The mutated amino acids are indicated in bold. Wild-type (WT) Vα17 chain in association with wild-type or mutant TCR Vβ5.1 chains was retrovirally transduced in the CD4 transfected 58α⁻β⁻ T-cell hybridoma variant[29]. Hybridoma transfectants expressing wild-type Vβ5.1/Vα17 (as a control, left panel), and mutated Vβ5.1(E→Q)/Vα17 (middle panel) or Vβ5.1(D→N)/Vα17 (right panel) TCR heterodimers were stimulated with either Q (filled diamonds) or E (open circles) peptide at the indicated concentrations. The IL-2 release values are the mean of triplicate cultures ± s.d. The data are representative of three independent experiments. **b**, A representative Q = E hybridoma was stimulated with Q (left panel) and E (right panel) peptides in the presence of wild-type HLA-DQ8 (WT DQ8, filled diamonds) or mutant HLA-DQ8 (D57 DQ8, open circles) with an Ala to Asp substitution at position β57. The level of IL-2 was measured using mouse cytotoxic T lymphocyte lines (CTLL) bioassay. The mean of duplicate cultures are shown, and data are representative of three separate experiments.

Table 2 | Negative charge in the CDR3β loop of human native gluten reactive T-cell clones

Clone	Specificity	Vβ	Jβ	Length	CDR3β	
Spec16 (FJ410946)	Q	7.1	J2–3	10	CAS SQ**d**irNTGEL	FFGEG
Spec12 (FJ410947)	Q=E	9	J1–5	7	CAS SQ**dd**QPQ	HFGDG
Spec12 (FJ410948)	Q=E	1	J2–5	10	CAS SVydgRGETQ	YFGPG
Spec13 (FJ410949)	Q=E	9	J1–5	7	CAS SQ**dd**QPQ	HFGDG
Spec13 (FJ410950)	Q=E	1	J2–5	11	CAS sttPGTgTETQ	YFGPG
TCC360.11 (FJ410951)	E>Q	3	J1–6	7	CGs sv**d**aSPL	HFGNG
TCC360.8 (FJ410952)	E>Q	1	J1–1	9	CAS SV**e**amNTEA	FFGQG
TCC489.2.1.4 (FJ410953)	E≫Q	1	J2–3	11	CAS SV**e**aGvsTDTQ	YFGPG
Spec22 (FJ410954)	Non-gluten	22	J1–5	10	CAS SlgpgdNQPQ	HFGDG
Spec1 (FJ410955)	Non-gluten	1	J2–3	11	CAS SVAGGlfTDTQ	YFGPG
TCC360.10 (FJ410956)	Non-gluten	13	J1–6	8	CAS reGyQETQ	YFGPG

DR3β sequences of gluten-specific HLA-DQ8-restricted and non-gluten-reactive control T-cell clones are presented. T-cell clones were obtained from research centres in Leiden (Spec16, 12, 13, 22 and 1) and Oslo (TCC360.11, TCC360.8, TCC489.2.1.4 and TCC360.10). Spec12 and Spec13 expressed two functional Vβ chains, which can be explained by the absence of allelic exclusion. GenBank accession numbers are in parentheses. Junctionally encoded amino acids are presented in lower-case. Amino acid residues at position 3 are bold when they are negatively charged.

previous observations showing strong *in vivo* and *in vitro* T-cell responses against peptides for which no binding to MHC could be detected by direct binding assays[21,22]. Furthermore, our data suggest that the instability of complexes between native peptide and HLA-DQ8 is compensated by the presence of a negatively charged residue in the CDR3β loop. In agreement with these findings, the presence of T cells during the short incubation time (1 h) of antigen-presenting cells with gluten peptides influenced the magnitude of their response against the native but not the deamidated peptide (Supplementary Fig. 3c).

Because the native gluten peptides recruit a T-cell population that is not only largely crossreactive but also substantially heteroclitic against the deamidated peptides, these cells could be recruited and continuously activated once inflammation is triggered and TG2 is activated to produce deamidated peptide. Consistent with this model, quantification of the CD4$^+$ T-cell response by enzyme-linked immunospot (ELISPOT) showed that the frequency of T cells recognizing the deamidated peptide was consistently highly increased after immunization with a mixture of native and deamidated peptides (Fig. 3).

To assess the pathogenic relevance of our findings further, we examined the TCR structure of human T-cell clones specific for the native DQ8-α–I epitope (QGSFQPSQQ). Notably, six out of six human T-cell clones, obtained from two independent research centres in Leiden and Oslo, expressed a TCR with a non-germline-encoded acidic residue precisely at position 3 of their CDR3β (Table 2). All these negatively charged residues were non-germline encoded and resulted from one non-germline (16.7%), two non-germline (33.3%), two palindromic (33.3%) and three non-germline (16.7%) nucleotide additions (Supplementary Table 3). In contrast, none of the three non gluten-specific T-cell clones derived from the same patients had this signature charge in their CDR3β (Table 3). To integrate these results in a more general context, we analysed 125 published

Figure 3 | Coimmunization with native and deamidated peptides amplifies the T-cell response to deamidated peptide. Humanized HLA-DQ8 transgenic mice were immunized with Q or E gluten peptides alone, or in combination. Eight days after immunization CD4$^+$ T cells were purified and re-challenged *in vitro* with 32.3 μM Q or E peptides. The frequency of IFN-γ-producing cells was assessed using ELISPOT. Each bar represents the mean spot number of triplicates ± s.d. **$P < 0.001$; data are representative of three independent experiments, each comprising three mice for each type of immunization.

TCR sequences obtained from HLA-DQ8 individuals with type I diabetes ($n = 116$), Birch ($n = 5$) and nickel ($n = 1$) allergy, or HIV infection ($n = 3$) and found that only 17 had a negatively charged residue at position 3 of the CDR3β loop. Interestingly, all of these 17 TCRs originated from the type I diabetes patients (data not shown).

This study reveals a new mechanism of peptide recognition in the context of HLA-DQ8, the MHC class II allele that predisposes to both type I diabetes and coeliac disease. The P9 pocket of HLA-DQ8, lacking the canonical Asp 57 found in other alleles, has the potential to interact not only with peptides harbouring an acidic residue at the P9 position, but also with TCRs harbouring a negative charge in their CDR3β loop. The potent response generated *in vivo* against native peptides with extremely low affinity for HLA-DQ8 may be explained by the ability of these TCRs to stabilize pMHC complexes. A large proportion of the TCRs recruited by native peptides display a crossreactive heteroclitic response to the deamidated peptide, which binds strongly to HLA-DQ8. Furthermore, because deamidated peptides do not impose charge constraints on the TCR, they can recruit an additional TCR repertoire. Consequently, the magnitude of the response against the deamidated peptide is increased under conditions in which native and deamidated peptides are both present. This may be particularly relevant in the context of coeliac disease, because TG2 is probably unevenly activated once inflammation is initiated. Furthermore, the magnitude of the T-cell response is viewed as a limiting factor for disease development[23,24]. In the absence of a disease model of coeliac disease, it is at present not possible to assess whether this process is important for disease onset, disease progression or other aspects of the disease. It also remains to be determined whether DQ β57 polymorphism also contributes to the strong association of coeliac disease with HLA-DQ2.

It is tempting to speculate that similar mechanisms may explain the association of type I diabetes with HLA-DQ8. The TCR repertoire of early pancreatic islet infiltrate is also characterized by the frequent presence of a negative charge at position 3 of CDR3β[19,20,25]. Furthermore, our preliminary studies suggest that TG2 activity is significantly increased in the pancreas, but not in the draining pancreatic lymph node and the spleen of 12-week-old NOD mice that have ongoing pancreatic inflammation (B.J. and L.T., unpublished observations). The ability to present native and deamidated peptides and recruit crossreactive TCR repertoires directly linked to the polymorphism at position β57 may underlie the remarkable association of HLA-DQ8 with severe immunopathology in mouse and human.

METHODS SUMMARY

Detailed methods are found in the Supplementary Information. Humanized HLA-DQ8 transgenic mice[15] were immunized in the footpad with immunodominant native (Q) and deamidated (E) α2-219–242 gluten peptide versions of this peptide with glutamates in position 229 and/or 237. T-cell proliferation, cytokine enzyme-linked immunoabsorbent (ELISA) and IFN-γ ELISPOT assays were performed to analyse the immune response to native and deamidated peptides. T-cell hybridomas were generated after immunization with native and deamidated gluten peptides and their specificity was characterized by
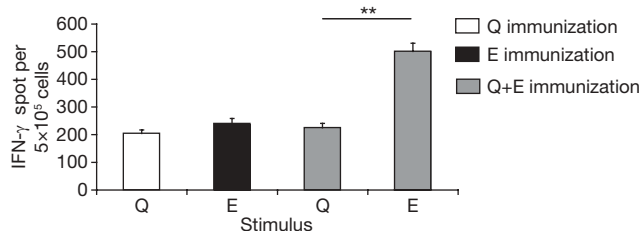
measuring IL-2 release[26]. TCR CDR3β sequencing was performed after TCR β-chain amplification with a set of TCR variable region (Vβ) sense and a TCR constant region (Cβ) antisense consensus primers as described[27]. TCR β-chain sequences were identified as TCR gene segments using BLAST and aligned/analysed using International Immunogenetics Database (http://imgt.cines.fr). The nomenclature is according to the World Health Organization-International Union of Immunological Societies (IUIS) nomenclature[28]. Transfectants expressing wild-type and mutated Vα17 and Vβ5.1 TCR of hybridomas that responded equally well to native and deamidated peptides were generated by retroviral transduction of T-cell hybridoma variant 58α⁻β⁻ (ref. 29). The soluble versions of wild-type and mutant DQ8 molecules were produced as previously described[30]. Human T-cell clones specific for the α2-219–242 epitope were established from intestinal biopsies of DQ8-positive coeliac disease patients as previously described[10,13].

1. Sollid, L. M. *et al.* Evidence for a primary association of celiac disease to a particular HLA-DQ α/β heterodimer. *J. Exp. Med.* **169**, 345–350 (1989).
2. Spurkland, A., Sollid, L. M., Polanco, I., Vartdal, F. & Thorsby, E. HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum. Immunol.* **35**, 188–192 (1992).
3. Acha-Orbea, H. & McDevitt, H. O. The first domain of the nonobese diabetic mouse class II I-A β chain is unique. *Proc. Natl Acad. Sci. USA* **84**, 2435–2439 (1987).
4. Hattori, M. *et al.* The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. *Science* **231**, 733–735 (1986).
5. Molberg, O. *et al.* Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nature Med.* **4**, 713–717 (1998).
6. van de Wal, Y. *et al.* Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J. Immunol.* **161**, 1585–1588 (1998).
7. Vader, L. W. *et al.* Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J. Exp. Med.* **195**, 643–649 (2002).
8. Siegel, M. *et al.* Extracellular transglutaminase 2 is catalytically inactive, but is transiently activated upon tissue injury. *PLoS ONE* **3**, e1861 (2008).
9. Henderson, K. N. *et al.* A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* **27**, 23–34 (2007).
10. van de Wal, Y. *et al.* Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc. Natl Acad. Sci. USA* **95**, 10050–10054 (1998).
11. Vader, W. *et al.* The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* **122**, 1729–1737 (2002).
12. Jabri, B. & Sollid, L. M. Mechanisms of disease: immunopathogenesis of celiac disease. *Nature Clin. Pract. Gastroenterol. Hepatol.* **3**, 516–525 (2006).
13. Lundin, K. E. *et al.* Gliadin-specific, HLA-DQ (α1*0501, β1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J. Exp. Med.* **178**, 187–196 (1993).
14. Lundin, K. E., Scott, H., Fausa, O., Thorsby, E. & Sollid, L. M. T cells from the small intestinal mucosa of a FR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. *Hum. Immunol.* **41**, 285–291 (1994).
15. Black, K. E., Murray, J. A. & David, C. S. HLA-DQ determines the response to exogenous wheat proteins: a model of gluten sensitivity in transgenic knockout mice. *J. Immunol.* **169**, 5595–5600 (2002).
16. Moustakas, A. K. *et al.* Structure of celiac disease-associated HLA-DQ8 and non-associated HLA-DQ9 alleles in complex with two disease-specific epitopes. *Int. Immunol.* **12**, 1157–1166 (2000).
17. Solinger, A. M., Ultee, M. E., Margoliash, E. & Schwartz, R. H. T-lymphocyte response to cytochrome c. I. Demonstration of a T-cell heteroclitic proliferative response and identification of a topographic antigenic determinant on pigeon cytochrome c whose immune recognition requires two complementing major histocompatibility complex-linked immune response genes. *J. Exp. Med.* **150**, 830–848 (1979).
18. Bankovich, A. J., Girvin, A. T., Moesta, A. K. & Garcia, K. C. Peptide register shifting within the MHC groove: theory becomes reality. *Mol. Immunol.* **40**, 1033–1039 (2004).
19. Baker, F. J., Lee, M., Chien, Y. H. & Davis, M. M. Restricted islet-cell reactive T cell repertoire of early pancreatic islet infiltrates in NOD mice. *Proc. Natl Acad. Sci. USA* **99**, 9374–9379 (2002).
20. Yang, Y., Charlton, B., Shimada, A., Dal Canto, R. & Fathman, C. G. Monoclonal T cells identified in early NOD islet infiltrates. *Immunity* **4**, 189–194 (1996).
21. Chao, C. C., Sytwu, H. K., Chen, E. L., Toma, J. & McDevitt, H. O. The role of MHC class II molecules in susceptibility to type I diabetes: identification of peptide epitopes and characterization of the T cell repertoire. *Proc. Natl Acad. Sci. USA* **96**, 9299–9304 (1999).
22. Yu, B., Gauthier, L., Hausmann, D. H. & Wucherpfennig, K. W. Binding of conserved islet peptides by human and murine MHC class II molecules associated with susceptibility to type I diabetes. *Eur. J. Immunol.* **30**, 2497–2506 (2000).
23. Ploski, R., Ek, J., Thorsby, E. & Sollid, L. M. On the HLA-DQ (α1*0501, β1*0201)-associated susceptibility in celiac disease: a possible gene dosage effect of DQB1*0201. *Tissue Antigens* **41**, 173–177 (1993).
24. Vader, W. *et al.* The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc. Natl Acad. Sci. USA* **100**, 12390–12395 (2003).
25. Quinn, A. *et al.* T cells to a dominant epitope of GAD65 express a public CDR3 motif. *Int. Immunol.* **18**, 967–979 (2006).
26. Bendelac, A. *et al.* CD1 recognition by mouse NK1⁺ T lymphocytes. *Science* **268**, 863–865 (1995).
27. Casanova, J. L., Romero, P., Widmann, C., Kourilsky, P. & Maryanski, J. L. T cell receptor genes in a series of class I major histocompatibility complex-restricted cytotoxic T lymphocyte clones specific for a *Plasmodium berghei* nonapeptide: implications for T cell allelic exclusion and antigen-specific repertoire. *J. Exp. Med.* **174**, 1371–1383 (1991).
28. Arden, B., Clark, S. P., Kabelitz, D. & Mak, T. W. Mouse T cell receptor variable gene segment families. *Immunogenetics* **42**, 501–530 (1995).
29. Blank, U., Boitel, B., Mege, D., Ermonval, M. & Acuto, O. Analysis of tetanus toxin peptide/DR recognition by human T cell receptors reconstituted into a murine T cell hybridoma. *Eur. J. Immunol.* **23**, 3057–3065 (1993).
30. Scott, C. A., Garcia, K. C., Carbone, F. R., Wilson, I. A. & Teyton, L. Role of chain pairing for the production of functional soluble IA major histocompatibility complex class II molecules. *J. Exp. Med.* **183**, 2087–2095 (1996).

naturejobs

"Calm down. All you have to do is write a thousand words and everything will be fine. And you have all day, except it's already noon." Thus begins *Violet*, a text-based interactive computer game that has the player take on the role of a stressed-out graduate student with a dissertation deadline and a beleaguered beau who is threatening to flee the relationship — and the continent.

Violet is the academic equivalent of *Zork*, the first popular work of interactive fiction, written by a group of Massachusetts Institute of Technology computer scientists from 1977 to 1979. *Zork* was considered both an amazing innovation and a massive time-waster. The same could be said for *Violet*, which won the 14th annual interactive fiction competition. *Zork* was deemed innovative because it created a fictitious world that a player could navigate by typing in sentences. *Violet*, the creation of Jeremy Freese, a professor of sociology at Northwestern University in Evanston, Illinois, is unusual in that the player takes on a mundane role, plagued by banal real-world irritations.

So, does *Violet* have any utility? One commentator on an education-technology blog sponsored by *The Chronicle of Higher Education* (http://chronicle.com) doesn't see the appeal. "What a redundant concept. Video games are a distraction from dissertations and research, now there is a video game about our dissertations and research. Is there an option to play a video game to escape?"

Maybe the best audience for *Violet* isn't graduate students with their own dissertation deadlines, but prospective students, who can play the game to help them decide if they really want to 'go there'. *Violet* could also be required playing for administrators, department heads, university presidents and deans, as both a reality check and a sort of graduate-student empathy builder.

Perhaps one change is required to make the game more widely accessible — something that honours its interactive predecessor as well as its affectionately geeky potential players. One name comes easily to mind: Dork …

Paul Smaglik moderates the *Naturejobs* Nature Network career-advice forum.

SEA OF DREAMS

They may have first been lured by the romance of underwater discovery, but oceanographers are now riding a wave of career growth and recognition, says **Quirin Schiermeier**.

M. JONES/NOCS

As a boy, Martin Visbeck dreamt of becoming a ship's captain. He soon realized that modern seafaring lacked the romantic adventure he aspired to, but his love for the oceans endured. After initially studying meteorology, he soon switched to oceanography, earning his PhD in 1993 from the University of Kiel in Germany. Now, as chairman of the physical oceanography research unit at the Leibniz Institute of Marine Sciences in Kiel, he has been a chief scientist on several research cruises, still striving to live out that boyhood dream.

Such career tracks, starting in fields such as physics or meteorology, are not unusual in oceanography. Although the oceans have been the subject of intense scientific interest for centuries, oceanography gained recognition as an independent scientific discipline only a couple of decades ago. In recent years, oceanographers' skills and expertise have been increasingly sought after.

"The global economy needs us," says Rachel Mills, a senior researcher at the National Oceanography Centre (NOCS) in Southampton, UK. "From the science of climate change to technical solutions for carbon-dioxide sequestration, and from oil exploration to maritime law and ocean policy, there are many areas where oceanographers' skills are being requested."

To the general public, oceanographers have an image as explorers of a fascinating, alien underwater landscape. In 1977, for example, they discovered hydrothermal vents in the Pacific that sustained bizarre life forms. This was hailed as a spectacular triumph and the findings became a popular topic of science books and documentaries. Oceanographic achievements have as yet never piqued the interest of Nobel prize committees and Visbeck fears they will not do so any time soon. Yet these stereotypes of a romantic field of underwater exploration belie the rigorousness of the science.

Take physical oceanography, for example: studies of oceanic motions, from small-scale mixing processes to basin-wide circulation patterns. This requires a thorough understanding of fluid mechanics and the laws of thermodynamics. Likewise, ocean biogeochemistry — studies of the way carbon and nutrient cycles, ocean chemistry, food chains and ecosystems respond to global warming — requires a great deal of expertise in molecular biology and organic and inorganic chemistry.

Physicists and engineers welcome

"People often think that what we're doing is all about cruising, diving and watching animals," says Visbeck. "Students are then often surprised to find out how much physics and mathematics we actually use."

Ocean modellers — whose work involves modelling flows, internal tides, mixing processes and ocean circulation — often prefer mathematically inclined PhD students with a physics or engineering background to students trained in less formalized disciplines such as environmental sciences or geography.

Since opening in 1998, the NOCS School of Ocean and Earth Science has produced more than 250 PhDs, the vast majority of whom have found jobs, according to the centre. The largest group (37%) has stayed in academic research, about half still in Britain.

"People knew extraordinarily little about the oceans until just a few decades ago, so it's a field where you can really still make a contribution," says Mills. "People thought I was mad when I left chemistry in favour of oceanography, but I found this new field just so much more exciting."

Hendrik van Aken, a physical oceanographer at the Royal Netherlands Institute for Sea Research on the Friesian island of Texel, says his teachers were sceptical

C. CHARETTE

"They have such good quantitative and analytical skills, they can find jobs in many areas outside academic science."

— James Yoder

when he entered the field in the 1970s. "All my professors would tell me, 'Don't do it, there are no jobs,'" he says.

Their fears were groundless. Throughout his academic career, van Aken has seen his research grow in importance to environmental and coastal planners in the Netherlands. He investigates the characteristics of silt transport into the Wadden Sea from the coasts of the Dutch North Sea. Such studies of sediment transport, initially intended as basic research, often prove valuable for more practical projects — for example, the ongoing Rotterdam harbour enlargement. And the relevance of oceanographic work to exploration and engineering activities has created research opportunities. Over the years, many of his former students and postdocs have been recruited into coastal engineering and environmental-management departments, in the Netherlands and around the world, he says.

Dip your toes in the water

During the past 20 years, for example, van Aken's former student Edward Kleverlaan has dabbled in a multitude of oceanography areas. His career path epitomizes the potential breadth of opportunities afforded to those with oceanography training. After receiving his master's degree in oceanography, he was recruited by Shell UK to do surveys in the North Sea; he then returned to the Netherlands to coordinate coastal protection projects funded by the Dutch ministry of transport and water management. In 1990 he moved to Australia, where he worked as a scientific research officer with the government of Queensland, and later with the federal government of Australia as an expert on climate change, water quality and marine conservation. He was also involved in managing the Great Barrier Reef Marine Park.

Kleverlaan, now a technical officer with the United Nations' International Maritime Organization (IMO) in London, has met many young oceanographers in various professions. "They're with consulting firms, with the geological surveys, in drilling technology development, mapping, remote sensing — just everywhere," he says.

As the importance of oceans grows in political and environmental arenas, many more oceanographers are likely to pursue careers at government agencies and international organizations, says Kleverlaan. And it's not only obvious employers such as the IMO that need them. About a third of people with doctorates from the NOCS find employment at government agencies such as the official British weather forecaster, the Met Office.

Most oceanographers who leave academia remain committed to careers in the environmental sector. However, the biggest job boom has been sparked by growing demand in the oil-exploration industry. Companies such as Shell have in recent years been desperate for exploration experts willing to spend extended time on drilling ships, says Mills. "It's definitely not everybody's taste, but at 20 to 30 many are still flexible enough and happy to move around," she says. A few years of ship-based work — whether assisting with test drilling, exploration or seismic studies — can be a stepping stone to other jobs in the business; field- or office-based, up to senior management.

However, in some regions, keen interest in academic careers has contributed to a bottleneck. Even more than in Europe, young oceanographers trained at US institutions tend to think of academia first when it

NOCS students work on the RRS James Cook (left, in Antigua) and the RV Callista (above).

M. COOPER

comes to career planning. This is causing concern at the Woods Hole Oceanographic Institution (WHOI) on Cape Cod, Massachusetts, says James Yoder, vice-president for academic programmes. The WHOI operates a prestigious marine-sciences graduate programme in partnership with the Massachusetts Institute of Technology in Cambridge; it hosts some 70 postdocs at any time. Interest has notably increased during the past decade or so, says Yoder, but the number of permanent positions available at US universities and oceanographic institutions has not.

The uncertain economic outlook adds to fears of an increasingly tight bottleneck. The WHOI, although a private corporation, depends on federal research dollars. The same is true of the Scripps Institution of Oceanography in La Jolla, California, and university departments in places such as Rhode Island, Florida, Texas and Oregon. The US National Science Foundation has already cut back its Antarctic science programme because of high fuel prices. During his campaign, president-elect Barack Obama pledged to increase science budgets — but it's still early days.

Regardless of how budgets will develop, Yoder encourages PhD students and postdocs to think more broadly. At the WHOI's yearly career forums, invited speakers from private companies, consulting firms, publishing houses and other sectors provide first-hand information about emerging job opportunities in their respective beats. "Oceanographers have such good quantitative and analytical skills, they can find jobs in many areas outside the narrow field of academic science," he says. "We just need to make them aware of the many opportunities that exist out there."

For marine scientists such as Visbeck, the allure of the oceans curtails any major concerns about funding and job prospects. Just 10% of oceanographers trained at the NOCS, for example, have in the past ten years opted for careers completely outside the field. With global energy in transition, niche areas such as tidal energy and offshore wind parks are likely to proliferate, offering opportunities to the next generation. "There may not be a fixed job outline for oceanographers," says Visbeck, "but it's certainly no longer necessary to discourage young scientists from choosing the field." ■

Quirin Schiermeier is *Nature's* **Germany correspondent.**

Buoyant prospects: Rachel Mills (top), Martin Visbeck.

D. TEAGLE/NOCS

MOVERS

Erich Nigg, director of the Biozentrum, University of Basel, Basel, Switzerland

1999–2008: Director, Department of Cell Biology, Max Planck Institute for Biochemistry, Martinsried, Germany
1995–99: Department of Molecular Biology, University of Geneva, Switzerland
1987–95: Swiss Institute for Experimental Cancer Research, Epalinges, Switzerland

After earning a PhD in biochemistry at the Swiss Federal Institute of Technology in Zurich (ETHZ) in 1976, Erich Nigg sought culturally and scientifically enriching experiences in California and Switzerland's French-speaking region before he was lured to Germany's Max Planck Institute for Biochemistry for the past decade. He is now headed home again to lead the University of Basel's Biozentrum.

Nigg's moves have prompted several shifts in his research focus. That eclectic background positioned him to head the Biozentrum, one of the first basic research institutes to foster interdisciplinary collaborations in cell biology, microbiology and structural biology. But, he says, "the price I've paid for changing my scientific research many times during my early career is the long time to gain recognition".

Nigg's initial focus was on cell-membrane dynamics and his doctoral dissertation explored protein mobility through cancer cell membranes. He followed this interest to sunny San Diego, to a postdoc position with John Singer at the University of California. "You are never as free to focus solely on your own work as you are when doing a postdoc," he says, regretting that some students don't take the opportunity to study abroad and learn about other cultures. At the University of California, he learned to use antibodies to visualize cell structures.

After returning to the ETZH for a few years, he opted for a tenure-track position at the Swiss Institute for Experimental Cancer Research, where he concentrated on the emerging topic of cell-cycle regulation by kinases. After a period at the University of Geneva, he was offered a directorship at the Max Planck Institute for Biochemistry in Martinsried, Germany. Nigg couldn't turn down the sheer power of Max Plank's instrumentation and financial support, which over the past decade has enabled him to study cell division using both cell biology and proteomics.

Nigg plans to use his varied experience to help young Biozentrum faculty members achieve their career goals. The Biozentrum had a common complaint — a generation gap had opened up when the original scientific leaders retired, says Susan Gasser, director of the Novartis-funded Friedrich Miescher Institute for Biomedical Research (FMI) in Basel. She helped convince Nigg that the region was developing into a strong biosciences hub. "Erich's recruitment demonstrates that Switzerland is able to attract back top scientists who took leadership positions in other countries," she says.
Virginia Gewin

BRICKS & MORTAR

Toyota motors ahead

Despite a significant downturn in the US car industry, Toyota is going ahead with plans to establish a research institute in Michigan. It will employ 60–70 scientists, who will focus largely on developing ways to mitigate motor vehicles' environmental impact.

Last April the company announced the creation of the Toyota Research Institute of North America in Ann Arbor. It plans to spend US$100 million during the next four years on research topics, to include energy efficiency, environmentally friendly materials and fuel optimization.

Toyota spokeswoman Cindy Mahalak says the shaky economy isn't affecting plans for the research institute. "The $100 million was not an additional investment. It was already in place," she says. The institute's use of an existing space in the Toyota Technical Center helps bolster the research budget, she adds. Leading the institute is Noboru Kikuchi, a professor of mechanical engineering at the University of Michigan and a director of Toyota research and development laboratories in Japan.

The institute has already hired 40 researchers and administrative staff, Mahalak says. It plans to add 10 more research scientists by the end of this year and another 20 by 2010. Toyota is currently advertising five research positions (most requiring a doctorate) in the areas of microelectric mechanical systems, biotech materials, nanomaterials and battery research.

The other positions will include opportunities for postdocs and students, according to a company spokesman. Earlier this year, Toyota advertised for two collaborators from academic universities and national labs for research in catalysis and nanotechnology at the centre as part of its quest to cut greenhouse-gas emissions. The spokesman declined to comment on whether the collaborations are in place.

According to one industry analyst, Toyota's decision to go ahead with its research institute plans in the face of a weak industry reflects the company's strong financial position globally. "The auto manufacturers are not all in the same position, by any means," says Tom Libby, senior director of industry analysis for J. D. Power and Associates in Westlake Village, California. Pursuing research even in an uncertain economic climate is a wise move, he adds. "You need new products and processes to remain competitive."
Karen Kaplan

POSTDOC JOURNAL

Calm before the storm

When I started, I wasn't sure what my new department was expecting of me. Would I immediately be asked to perform like the other permanent academic staff, deftly juggling teaching and research commitments while simultaneously balancing administrative responsibilities? Day one did begin with the inevitable admin, but with secretaries' help it was quickly done, causing minimal stress.

I braced myself for the second wave of work. Instead, I received more time — time I used to think about my research. I pinched myself, but the free time didn't go away. I had time to plan how I'm going to teach maths-phobic students about mathematical modelling and time to apply for funding. There might even have been time for a cup of tea, if our common room wasn't due for demolition.

I think this initial calm is the opportunity to set my research direction within the department for the next few years. I do not expect it to last. So I am using the time to foster local collaborations, submit funding proposals, advertise myself to seminar organizers and even plan the lectures that will eventually compete with my research for time and attention.

Perhaps I should feel guilty when I see Paul scurrying to another lecture, or Carl looking exhausted after a night of proofreading his student's thesis. But I don't. Because I know my time will come.
Jon Yearsley is a lecturer at University College Dublin's school of biology and environmental science.

From Mars with love

He who dares …

Julian Tang

"So children, what would you like to hear about today?"

The Old-Timer smiled down, benignly, at the 30-odd youngsters sitting cross-legged in front of him on the purple Martian grass.

"A story!" they yelled in unison, giggling and rocking about on their bottoms, knocking each other over.

The Old-Timer also started to laugh, but quickly stopped himself as he started to cough and splutter. With the weaker Martian gravity and greater distance from the Sun, the human colonists on Mars had increased their average life span to about 150 Earth years. Yet, even by these standards he was really old, being almost 200 Earth years now.

"OK, can anyone tell me when humans first came to Mars?"

There was a momentary silence then a few hands shot up. He pointed to a cute-looking girl dressed in a one-piece, silver dress.

"A very long time ago," she said with utmost seriousness and sincerity, looking up at him with big, round eyes.

"About the same time as you were born, Old-Timer!" a young boy, about six or seven years old, sitting beside her shouted out, cheekily. There followed some ill-concealed sniggers.

The Old-Timer looked at them, beaming and said: "Yes, you are both right! In the late twenty-first century, the first human finally managed to land on Mars. Scientists had managed to overcome the long journey time with some great advances in rocket technology, introduced by …" and he paused dramatically, "… by my father, then on Earth."

There was a hushed, respectful silence following this announcement.

"And does anyone know who the first astronaut to Mars was?" He looked around, hopefully.

Many of them shook their heads.

"Who was that, sir?" asked one of the older boys, finally, from the back row.

"Well, let's see now," said the Old-Timer appearing to think, carefully. "That first astronaut, well, that would have been me."

Another shocked silence followed.

"Really?" said the cute, big-eyed girl.

"Yes, really," replied the Old-Timer, looking at her kindly. "But there was a problem. After I returned, there must have been something different about me. My wife said that she did not know me. Even I did not know who I was."

"What happened then?" asked the cheeky boy, curiously.

"Well, eventually, after about six months, I finally remembered who I was, and everything was OK. Then, I had to train other astronauts about how to return to a normal life on Earth — after more than a year alone in space!" he finished abruptly, with grin.

Instead of laughter, there was an unexpected silence.

Then with a realization the Old-Timer looked up at the sky and chuckled to himself. "Ah, yes, well, I guess none of you have yet been away from your friends and parents for so long," he said, smiling down at them again.

Many heads shook vigorously, seemingly in horror at the thought of this. There was another uncomfortable silence. It seemed that their earlier cheeriness had disappeared.

The Old-Timer sighed, looking at his watch. He could only keep the old memories away for so long. "OK kids, it looks like I've given you something a bit extra to think about today and I'm feeling a bit tired now. I'll make it up to you tomorrow, OK?"

The children murmured their acceptance and moved off in small groups, talking quietly among themselves. The Old-Timer looked after them, smiling wistfully. Maybe there was a future explorer among them yet.

But had they seen him a minute later, they would have seen his smile turn into a grimace, perhaps of pain, and then … to something … not quite human.

His face had turned the colour of red Martian soil, seeming to continually shift and reshape itself, like windblown sand swirling across his facial landscape.

He slapped his hands over his distorted face, gasping through clenched teeth, "No! Leave them alone! You will not escape! Not now, not ever! You destroyed life here once before and will not do it again. As long as you are within me, I cannot die and you cannot escape. I even had to kill her to keep this secret."

JACEY

A sob of deep human grief escaped the dustiness of his lips. "My beautiful wife," he whispered. "For that, you and I shall suffer for all eternity."

With this, like a signal for a temporary ceasefire, his face gradually returned to normal, lined and old, but fully human once more.

He looked wearily towards the distant, setting Sun, still remembering like yesterday, landing on this bare Martian soil, proudly making the first human footprint, and how this had awoken something in the red sand. Something that had waited for eons … something that was old and hungry for life. It had passed through his suit without resistance, penetrating his skin, filling him up from the inside out. Yet, he had survived, contained and controlled it. He knew then that he had to return to Earth as if all was well, to stay alive and protect humanity. For if he had let himself die here, it would have escaped back into the red sand. Others would come after him, unsuspecting, to investigate. Others, who, unlike him, may not have been able to control it, from whom it would have spread to others, perhaps to every soul on Earth. It would have consumed them all, leaving Earth as it did Mars, a lifeless, desert wasteland.

This was his pain, his suffering, his penance, for his arrogance, his desire to be the first. Now, he must bear the burden, the curse of his achievement, until perhaps a future generation could relieve him of the monster that lived within him, and allow him, finally, to die in peace. ∎

Julian Tang is a Singapore-based virologist, who, among other things, enjoys reading science fiction, fantasy and horror. He has many story ideas, but needs time to commit them to paper.